

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 19:22:50 : Search time 2662 Seconds
(without alignments)
16815.125 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139
Sequence: 1 atgcaaatgaattaaaaaag.....aaagagctgaagtgtactctaa 2139

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlgo_inv:*

BEST AVAILABLE COPY

C	1	129.4	6.0	10689	1	AE004008	AE004008 Xylella f
C	2	43.2	2.0	125020	3	AF429315	AF429315 Homo sapi
C	3	42.2	2.0	1461	3	AY051790	AY051790 Drosophila
C	4	42.2	2.0	76748	2	AC014256	AC014256 Drosophila
C	5	42.2	2.0	148847	3	AC008234	AC008234 Drosophila
C	6	42.2	2.0	162593	3	AC008356	AC008356 Drosophila
C	7	42.2	2.0	224400	3	AE003682	AE003682 Drosophila
C	8	41.6	1.9	720	8	CNS01DRA	ALL16606 Botrytis
C	9	41.1	1.9	1700	1	RHMTBRC	M15810 R.mellioli
C	10	41.1	1.9	329100	1	SMES91787	AL591787 Sinorhizo
C	11	40.4	1.9	125020	9	AF429315	AF429315 Homo sapi
C	12	40.0	1.9	4314	5	AF051784	AF051784 Xenopus l
C	13	39.4	1.8	3060	8	EMERLBA	L24395 Emericella
C	14	38.8	1.8	348450	1	MLBPRM4	AL583920 Mycobacte
C	15	38.6	1.8	10624	1	PDSEX	X79242 Paracoccus
C	16	38.6	1.8	156971	2	AC099055	AC099055 Homo sapi
C	17	38.4	1.8	100306	2	AC105363	AC105363 Oryza sat
C	18	38.2	1.8	190050	2	AL646080	AL646080 Ralstonia
C	19	37.8	1.8	31896	2	AC017879	AC017879 Drosophila
C	20	37.8	1.8	172904	3	AC007414	AC007414 Drosophila
C	21	37.8	1.8	260910	3	AE003831	AE003831 Drosophila
C	22	37.4	1.7	38408	2	AC019786	AC019786 Drosophila
C	23	37.4	1.7	146153	3	AC007452	AC007452 Drosophila
C	24	37.4	1.7	203356	2	AL669952	AL669952 Mus muscu
C	25	37.4	1.7	264646	3	AE003818	AE003818 Drosophila
C	26	37.2	1.7	133691	9	AC074347	AC074347 Homo sapi
C	27	37.2	1.7	170398	2	AC013540	AC013540 Homo sapi
C	28	37.2	1.7	329709	1	AP002997	AP002997 Mesorhizo
C	29	36.8	1.7	810	6	AX110931	AX110931 Sequence
C	30	36.8	1.7	2016	8	PAN296102	AJ296102 Podospira
C	31	36.8	1.7	10925	2	AE004750	AE004750 Pseudomon
C	32	36.8	1.7	128135	2	AC104321	AC104321 Oryza sat
C	33	36.8	1.7	307150	1	CNSPAX01	AJ248283 Pyrococcus
C	34	36.8	1.7	349980	6	AX041106	AX041106 Sequence
C	35	36.6	1.7	1129	8	AF178952	AF178952 Zea mays
C	36	36.6	1.7	1496	8	CRENOLM	X66412 C.reinhardt
C	37	36.6	1.7	6709	14	MOCDRNAP	L38927 Molluscum c
C	38	36.6	1.7	170630	2	AC016905	AC016905 Homo sapi
C	39	36.6	1.7	185097	2	AC023148	AC023148 Homo sapi
C	40	36.6	1.7	190289	14	MC060315	U60315 Molluscum c
C	41	36.6	1.7	212956	2	AC090543	AC090543 Homo sapi
C	42	36.6	1.7	222759	2	AC073177	AC073177 Homo sapi
C	43	36.4	1.7	858	9	AF197129	AF197129 Homo sapi
C	44	36.4	1.7	1091	9	AY064416	AY064416 Homo sapi
C	45	36.4	1.7	1503	9	AK000576	AK000576 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AE004008/c
DEFINITION Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION AE004008
VERSION AE004008.1 GI:9106961
KEYWORDS
SOURCE
ORGANISM Xylella fastidiosa 9a5c.
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.
REFERENCE
AUTHORS 1 (bases 1 to 10689)
Simpson,A.U., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M., Aya,J.E., Bala,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Boye,J.M., Britones,M.R.,
Bueno,M.R., Camargo,A.A., Camargo,L.E., Carro,D.M., Carreir,H.,
Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.B., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorrry,H.,
Facincani,A.P., Ferreira,A.U., Ferreira,V.C., Ferro,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Fritume,M., Furian,L.R.,
Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hohnsels,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.
and Marino,C.L.

Result Query Match Length DB ID Description

TITLE The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis
JOURNAL Nature 406 (6792), 151-157 (2000)
PUBMED 20365717
AUTHORS 2 (bases 1 to 10689)
Simpson,A.J.G., Rehnach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S., Barros,M.R., Bonaccorsi,E.D., Bordin,S., Boye,J.M., Brites,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carter,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cistofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facincan,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franco,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hohelsel,J.D., Jungelira,M.L., Kemper,E.L., KitaJima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Martins,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.L., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pesquero,J.B., Quaggio,B.R., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Melandris,J. and Setubal,J.C.
Direct Submission
TITLE Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil
FEATURES
source Location/Qualifiers
1. 10689
/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"
/clone="9a5c"
gene complement(119. 367)
/gene="XF1871"
CDS complement(119. 367)
/gene="XF1871"
/note="hypothetical protein: identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84677.1"
/db_xref="GI:9106962"
/translation="MTMTLKEDKTEIIDEQRIQAKLISEAKIQAEIHEVNAHQKL MKETLKLTESKWPVYVSSSLIAGATAATLFLFH"
gene complement(430. 618)
/gene="XF1872"
CDS complement(430. 618)
/gene="XF1872"
/note="hypothetical protein: identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84678.1"
/db_xref="GI:9106963"
/translation="MMEBARATAPQGGIASDPRTDAPKKTGLRKRKPYPTGSASIVLRHVLTORQOKRKASP"
gene 941. 1288
/gene="XF1873"
CDS 941. 1288
/gene="XF1873"
/note="similar to GI1479340 (percent identity: 36 %/query

alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84679.1"
/db_xref="GI:9106964"
/translation="MSKNKINQIITGMAGFQFNGKLVTPERTIEQOLRMLST CTIAREQKMEBARATAPQGGIASDPQRTITDALKTVSHKRRRAAETASIIQLRQ VLIQROOKRKASP"
gene 1351. 1662
/gene="XF1874"
CDS 1351. 1662
/gene="XF1874"
/note="hypothetical protein: identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84680.1"
/db_xref="GI:9106965"
/translation="MTMTLKEDKTEIIDEQRIQAKLISEAKIQAEIHEVTHQEM NAHINEMARTQEMNARTQKLIETQKSKESQWYPVVASGLAAGATAATLFLKLF H"
gene complement(1839. 2918)
/gene="XF1875"
CDS complement(1839. 2918)
/gene="XF1875"
/note="similar to SP|P03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84681.1"
/db_xref="GI:9106966"
/translation="MLNLVTGAPGNGKTLVAVDWLRIQIEIDKSLVKSAGVPRSYTD IEGPVEAVRRLTGTVQVQSDAPEDMTTPGQSVIYDEAHMFPAGRPGRSDPPRYDL DTHRGGIDLNFVYQWPTKIHHLRLNGEYHILNRANGLOTAGIKRRSRQDDPYDI HOREKAEVEVWKFPKDRYALYASSTLFVSHKFRIPKRVMSALSCVYIGLVFHW YSPAHLEASVAGSAGASGASRLAASPLSSSRSLVSGMRTYAVLETESAPTLSCGV SDRSCRCFNTDGYOIDMSVVECRILASPLFPVHYHAYVSSSSAAAVSSPSSPS SSVSLSSVSASAPLIGSSVPSH"
gene complement(2918. 3202)
/gene="XF1876"
CDS complement(2918. 3202)
/gene="XF1876"
/note="similar to SP|P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84682.1"
/db_xref="GI:9106967"
/translation="MFSWMLMLRNILQGYVARGLVAGLALVYTPPLIVTSALNL IVSKMSGISADVLNIALMGFGEALSTIGSAMLRALQSLHGVKATP"
gene complement(3195. 4412)
/gene="XF1877"
CDS complement(3195. 4412)
/gene="XF1877"
/note="hypothetical protein: identified by sequence similarity: putative; ORF located using Glimmer/RBSfinder"
codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84683.1"
/db_xref="GI:9106968"
/translation="WCSRSFLLYFISALGLTVISFPFCLFQANVPPTEVVPYTSVRS SVSISDGVSTATFEARSTALVNGVRYTVPDVDISASTIGSLAKAVAVRGAFAFYVYS

```

MSGLINGAWVIDELTHEVMSPALKEIPVGTAVAMMMORPDGHVEYVTPDGLIAP
INAYOSLSPPOLIVRTSSGSPSSVSSSRNLYHLEGGGEYGHLSKIDQASPDVSSGLP
SVIILPDQLIVRSDPSTRYNANVLTPDSORCAVILLETYSALKLRSLDEDLKASAP
DOQPSGGASPPSSSGTAMPSFCSMASVYVCFIDMWSDEFLKPLVLPDVPYDKL
PQAKTWSGLGEGACPSPTTPIEFSGKTSVEISYOPCEDFALMPVIVIAITLA
AVIAGGFRGVKNV"
/translation="KMSYIDRGIAFFSLSTSLVEADNADAGSEPDAGAAVTALGAS
GAVLIGAKLIKLPALISVGMKWLKGAIFG"
complement(4724..4924)
/gene="XP1879"
complement(4481..4702)
/gene="XP1878"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084684.1"
/db_xref="GI:9106969"
/translation="MKSVIDRGIAFFSLSTSLVEADNADAGSEPDAGAAVTALGAS
GAVLIGAKLIKLPALISVGMKWLKGAIFG"
complement(4724..4924)
/gene="XP1879"
complement(4724..4924)
/gene="XP1879"
complement(4724..4924)
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA084685.1"
/db_xref="GI:9106970"
/translation="MAOOLMSLYCTQYDVEARTCSQAMVPPSLPISYEDVRIIL
PILVMOFLVAMEFHLFTVYVRD"
complement(4931..5236)
gene
Query Match 6.0%; Score 129.4; DB 1; Length 10689;
Best Local Similarity 44.9%; Pred. No. 1,1e-22;
Matches 888; Conservative 0; Mismatches 1051; Indels 38; Gaps 9;

```

```

722 AATACAGCAAGCAGCAATTAACCCCTATTAAGCCGCTTACTGCTGCCGTATCATGCAAG 781
9787 GCTTCTCTTAAGAAACATACCTACCGCTCCAAACACCTGCGTAAATTTAGCATCAAC 9128
782 GCTACAAAGCGCTGACGACTATGACATGACATCCGTTTCCCGGGCAGTACGATCGTACC 841
9727 CATTAGTGATGTGTGACTTCGTGATGTGCTGCTGTTTACCCAGACGATCAACATCGTTAG 9668
842 TCACCTCTTGGGGGTGGGAAGATGTAATGCAAAACAGAAACAAATCTGTATGAAATTC 901
9667 CTTTGGGCGCCGAATTTAGAAATACGTCTACGACTTATCCGGTATGTCGACAGATT 9608
902 GCGGTATCAAGCAAGCATGTGGAAGGAAGCCATGAGCCGAGATCAGGCTATCCCGATCA 961
9607 TCAAGAACTCATTTGCTTGTATTTAGGCGGCTGCAAAACGAATCCAGACATTCAGATGA 9548
962 AATATGCGCAGCAAGTATAGCTCAGAGTGCCTAATTTGGAAGAAATTCGATCGTATGAAAC 1021
9547 AATATGCGACCTTTATAGTGTGATTAATTAACATCGAATAAACTTTGATGTCATTTGG 9488
1022 GCGGTCTGCTGCTTGTGACGTATAGTGTGTAAGCGTCCGAGAAAGACATTCGACAG 1081
9487 ATGGCTTCCGACCATTAATGCTATAGTGTGCAAAACAGAGTGAAGAAACAGCATTAATAG 9428
1082 ACTGATCCCTGAAGAACGCAAGAGTGTCTATAGCCGATGATTTGCTCTCTCGAAA 1141
9427 CTTGTGTGAAGAACAGAGGGGATACGTGACACAGAAAGATTTGGTCCGATCAAAACATTGG 9368
1142 AGGCTTATTAAGAAAGAGCCAAAGCCAAACCGTATGATGATTTATTTGAGGAGAGACGCT 1201
9367 TTATCTGACAGAAACATACAAAGCAATACGATGCTGATTTTGTATTTGGTACAGTCA 9308
1202 TCG---GTGTTACGAGAGTGTGCTTTTTCACAGTTTGCCAAACGATTTGGTACAAATC 1258
9307 ACCGAAAGTGTGTGATTTGTTGAGCGGTGAACCTATTCCTTGGCAATGACGCGACAA 9248
1259 CTGATGCTCATGCGCGGATCTCAAAATCGCTTACAGCAAGTACAAAGATCTACCTCCCT 1318
9247 AATCTGATGCCCAACCGTGAAGCGGATATCAGAGCGCGATCTGCCCAACATTGAAGGCA 9188
1319 CGCTCGACCGCTAAGAGTGTCTCCGCAATGCTCATATTTAGCCCGGCGATCCCTGCGC 1378
9187 ACTTAAAGCAGATGTGAGCGACGTTACTTACCGGAATAGATTCGCCAG--ATGACGCAAT 9131
1379 ACAAGCTCCCGATATATTCAGAAATGTATGCAAA-----GAAATTCAAAGC 1428
9130 ACTGCTCACCGAATACAAACAGCTTCCTGCAAAACGCGTGGCTGCAATTGATGTTT 9071
1429 GACACGAAGAATATGACACTTCGTATTCGACAAAGAGTGTGCTTATACGACAAAG 1488
9070 GGTGGAGATGGCANTCCGCAACTCTTCAAGGCGTTGGGTGACACCCTTAATTAAG-CAGT 9012
1489 TTCCATGCGCATGCTCAAGTGCATGAGCAAGAAAGTTTCCCAAGGCTATTCGAAAGAT 1548
9011 TCTGAGGAAGAGCTGAGAGTGTATTAATGCCAGCGTACAGCTTTCGAACAGCTCAAGAT 8952
1549 CCGCAGTATAGGCTTTCCAAAGACGTAATAGCTGTCTGCTGCGCTATTTCAAGCGGATGCG 1608
8951 CCAGCGATCCGTTATGCTGTGGCCATCATCAGCAATTAAGGAGATTTGAACGTGAGAT 8892
1609 ATGGCCATGCTTATGCTATGAGAAAGGCAAGCGTCTTTCTTTTGGCGGTTGGCGTAG 1668
8891 AAAATCCGTACTGAGAACTACTTAAGCCGTCGATCTTTCGAAAGGCTTAAAGCTTAAAG 8832
1669 ATGTACCCCGGAGCTG-----CTCTGCGAGAGGATCCCACTTACATGATGCTATGAGC 1722
8831 TACACAGAGTATATGCAAAATTTGTTTACCCAGAGCCCACTCATCATATTACGATTTACC 8772
1723 TACGCTCATCAAGGATATGAACCGCAGAGCGTGCCTGTACAACTATTCATACGACA 1782
8771 TTGGCGCACGTAAGAAAGCTATTCACCCAAAGATGGCGTGAAGTACACACGTTTACGACA 8712

```

QY	1783	GGCAGGGCGATATGCAAGACGAGATCTTAAGACGATGATTTGGCCGTACGAGAGAT	1842
Db	8711	CTCAGAGGGTGTCAATGCGCA---AGATATACAGCGCTCGAGACCAFTTCGATTACCCGAATCA	8655
QY	1843	ATCCCTGCACCTCTCCGACACCAAAAATCATATGGTCCGTATGCGCA-----GAAACGTCAG	1896
Db	8654	CTGATTAATGGGATTAAGAGCAAAAGAGCTACGCCATTTAGCCGATCAGCGATTTGGCACT	8595
QY	1897	CTCCATATCGCTTTCCTCTATCGAACACATACACGGCGGTAATCCGGTATAGCCCCGTA	1956
Db	8594	GTCCCTGTGAATTTTCTATCTGACTTGACATTAATCGCGGCAACCTCGGATCGGCAGTCA	8535
QY	1957	TTTCGATGAAGACGGCGGCTCTATCGCTCTGCTTTCGATGATGGCACTGGGAACCTATGACT	2016
Db	8534	TTAGATATGCCACGGAAAAATTTGGTGGTTAGCATTTGACGGTAATTTGGAGTCTGTCACT	8475
QY	2017	GGTCACATCGAATTCGAACCCGATTCGACGCGCAACATACAGCGTGGACATCCGTACGTT	2076
Db	8474	AGTATTTGGGCTTTTGATCTCTGTATGACTCGAACATTTGCAGTTGATGATGCTTATGTG	8415
QY	2077	CTCTTCATGATTTGCAAAATGGGGTCAAGTCCCCGCTCATCCAAAGACGTGAAGTTG	2133
Db	8414	CAGTGATCATGACATCGAAGTTGCACACACACACATCATCTACTTAAGAAATTTGAATTTG	8358
RESULT 2	AF429315/c		
LOCUS	AF429315	125020 bp	DNA linear PRI 18-JAN-2002
DEFINITION	Homo sapiens junctophillin 3 (JPH3) gene, partial cds.		
ACCESSION	AF429315		
VERSION	AF429315.1	GI:17646244	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 125020)		
AUTHORS	Holmes,S.E., O'Hearen,E., Rosenblatt,A., Callahan,C., Hwang,H.S., Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A., Potter,N.T., Ross,C.A. and Margolis,R.L.		
TITLE	A repeat expansion in the gene encoding junctophillin-3 is associated with Huntington disease-like 2		
JOURNAL	Nat. Genet. 29 (4), 377-378 (2001)		
MEDLINE	21583737		
PUBMED	11694876		
REFERENCE	2 (bases 1 to 125020)		
AUTHORS	Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.		
TITLE	Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..125020		
source	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16q24.3: between D16S520 and WI-12410"		
	/note="Isolated from a patient with Huntington's disease-Like 2 (HDL2)"		
	complement(35581..35746)		
	/rpt_type="tandem		
	/rpt_unit="ctg		
	complement(<35507..>36887)		
	/gene="JPH3"		
	/product="Junctophillin 3"		
	complement(<35507..>36887)		
	/gene="JPH3"		
	/note="JP3"		
	complement(<35507..36887)		
	/gene="JPH3"		
	/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"		
	/codon_start=1		
	/product="Junctophillin 3"		

Query Match	2.0%	Score 43.2	DB 9	Length 125020
Best Local Similarity	9.8%	Pred. No. 5.2		
Matches 46	Conservative 226	Mismatches 194	Indels 4	Gaps 1
BASE COUNT	29056 a	32731 c	30696 g	28283 t 4254 others
ORIGIN				
OY	250	ATCTTTACCAACACCACCTGCGGATACGGTCTGATTCACAGACCCAAAGCAGCGGTGATGATC	309	
Db	51261	ACCMWYVYRYRYCKWMMMAAAMMCTAAAYYKWRMRWMSYSSVAYAMATAT	51202	
OY	310	GACATATCGCCGAGATGTTCTGTTCTTCGCACAGATGGGTGAGAGACTTCGATCCGGGT	369	
Db	51201	MCWYVYSSMYAARRRGYVSGGKKRMAWSTSCMCCRCCKKKKMGHRCSTTYTGMSCG	51142	
OY	370	CTTTCCTGGAAGATATCTCGCAAGATCTGTAAGGATTAACGCAACAAGTGAAGAAGACATC	429	
Db	51141	SRWMTSSMACMWMYKGSRRKTKSMATSCMBRGAIRSCKGRMKRYWSCRKGRMRGWSGY	51082	
OY	430	AAGGATACCTGACGACATGGAGCGTCTGCGCAAGCTCAGAGAGTATGCCAAGATG	489	
Db	51081	KGCSGYMGMWGSQRTSKYSYTGEGYCKSAKKCKSKTSKSCYTSKYS---CSMG	51026	
OY	490	GCCAAAAGAAAGATGACGAGACAGAACCACTCTCATCTGATAGACCTTCTTATTCAC	549	
Db	51025	KYSVCTCGWGRMCKGMSGCCASGCKSYTKMAGSMWCRGRGMSCHYWSMSKYSMAS	50966	
OY	550	AACGAATACCTTCATCATCTGATACGATATTCAGACAGCTTCGATGATTTCTCCT	609	
Db	50965	YKSSARGMKCYKTSCKMYYYSHMBKSHDBSGYKYSRYBWTCKMKMSKMSVDSH	50906	
OY	610	CCCAGCTCTGTAGTAATTCGCGAGCGCATACGCAACACTGGATGTGGCCCGTCACACG	669	
Db	50905	MWRDMYKMRKMYSSGRTGRYHBBAMWCKMSRSDSGYGMHMKYSKMSASRHWY	50846	
OY	670	GCGCATTACAGGATTCGCGGTATCTCGGATCGGCGGCAACCGCGCGC	719	
Db	50845	SGMKSSYVBMDCMSHTYBSKSVBRMRWSSGYBKRCAYMRKRIBHSS	50796	
RESULT 3				
LOCUS	AY051790	1461 bp	mRNA	linear
DEFINITION	Drosophila melanogaster LD30634 full length cDNA.			INV 27-AUG-2001
ACCESSION	AY051790			
VERSION	AY051790.1	GI:15291890		
KEYWORDS	FLI CDNA.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidae; Drosophilidae; Drosophila. 1 (bases 1 to 1461)			
AUTHORS	Stapleton,M., Brokstein,P., Hong,L., Abmayyan,A., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J., Nunoo,J., Pacliel,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M. and Celinker,S.			
TITLE	Direct Submission			
JOURNAL	Submitted (10-AUG-2001) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA			
COMMENT	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory Berkeley, CA 94720 This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene collection 1 (Rubin et al.,			

Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

source

1. 1461
Location/Qualifiers

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db_xref="taxon:7227"

/map="85D8-85D9"

1. 1461

/gene="CG8444"

/note="alignment with genomic scaffold AB003682"

/db_xref="FLYBASE:FBgn0037671"

53. 1015

/gene="CG8444"

/note="Longest ORF"

/codon_start=1

/db_xref="FLYBASE:FBgn0037671"

/product="LD30634p"

/protein_id="AAK93214.1"

/db_xref="GI:15291891"

/translation="MLRVVIFSLFIAINASGEPTVLRPKAIFSKNDALSHYVG
DLYVMGNVAVSGDPTNNGCTINDEPNIAKQITVIVHOGIGVTTAGNKKYTELGSG
TDSALNALAELEAANRPYCDINFEOPFDGVAMSCGDFEAPAPKPKHLPESLHT
ADKQFLQEVGFTNSADHIAEMAKPSNVLMLSVDAVAKAGEKSVAAVEENKLISA
AISRLAASOKSSDSVLEFVQTEKEDVAASRAKRDYIASTTPYMLAVYGSDDYVIE
NIILFMVVFGLIILCAIYALNAMPGRDSIIYRMTSPRIKRDN"

BASE COUNT 374 a 376 c 352 g 359 t

ORIGIN

Query Match

Best Local Similarity 49.3%; Pred. No. 6.6;

Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 1470 GGTTCCTTATGAGCAAGTTCATGCCATGCATGCATGACGAAAGAAAGTTTCG 1529

DB 703 GGCGGCTATCAGCGCGCTCTGCTGCTGCCAGAAAGTCCAGACTCGCTTCTGTG 762

QY 1530 CAAGCTATCGAAGAAATCCGAGTAGAGTTTCCAGAGCGTATAGCTGCTCTCG 1589

DB 763 CCAGACCGAGGAAAGAGAGCTGCGATCGCTGCCAACCGTGATACATTCCCGCTTC 822

QY 1590 CGCTATTGAGCGCGATGCGATGCGCAATGCTATGCAATGAGAGGCAAGCGTCTTT 1649

DB 823 CACACCAATCCCTACCAACCTGCGCTCTACTACGCGACGACTACCGGTATCTTCAA 882

QY 1650 CTTTGGCGGTTTCCGAGATGATACCCCGGAGCTGCTGCGCC 1692

DB 883 CATCATCTGTGTGTCATGCTCTTGGAGACTGCTCTGCTG 925

RESULT 4

AC014256/6 76748 bp DNA linear HTG 16-NOV-1999

LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered

DEFINITION pieces.

AC014256

AC014256.1 GI:6437079

KEYWORDS HTG; HTGS_PHASE2.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 76748)

Adams, M. and Venter, J.C.

Direct Submission

Submitted (16-NOV-1999)

Rockville, MD, USA

This sequence was identified as CDM:10211789 by the submitter.

For further information on this sequence e-mail to fly@celera.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

Location/Qualifiers

1. 76748

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 21897 a 15665 c 15830 g 23356 t

ORIGIN

Query Match 2.0%; Score 42.2; DB 2; Length 76748;

Best Local Similarity 49.3%; Pred. No. 9.2; 113; Indels 0; Gaps 0;

Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 1470 GGTTCCTTATGAGCAAGTTCATGCCATGCATGCATGACGAAAGAAAGTTTCG 1529

DB 17172 GGCGGCTATCAGCGCGCTCTGCTGCTGCCAGAAAGTCCAGAGACTCGTCTCTGTG 17113

QY 1530 CAAGCTATCGAAGAAATCCGAGTAGAGTTTCCAGAGCGTATAGCTGCTCTCG 1589

DB 17112 CCAGACCGAGGAAAGAGAGCTGCGATCGCTGCCAACCGTGATACATTCCCGCTTC 17053

QY 1590 CGCTATTGAGCGCGATGCGATGCGCAATGCTATGCAATGAGAGGCAAGCGTCTTT 1649

DB 17052 CACACCAATCCCTACCAACCTGCGCTCTACTACGCGAGCATACCCGGTGATCTTCAA 16993

QY 1650 CTTTGGCGGTTTCCGAGATGATACCCCGGAGCTGCTGCGCC 1692

DB 16992 CATCATCTGTGTGTCATGCTCTTGGAGACTGCTCTGCTG 16950

RESULT 5

AC008234 148847 bp DNA linear INV 17-FEB-2001

LOCUS Drosophila melanogaster, chromosome 3R, region 85D-85D, BAC clone

DEFINITION BAC32M04, complete sequence.

AC008234

AC008234.8 GI:12957667

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 148847)

Holt, R.A., Evans, C.A., Gocayne, J.D., Amaratilake, B., Tyler, D., Man, K.H.,

Rogers, Y., An, H., Baldwin, D., Bonzom, J., Beeson, K.Y., Busam, D.A.,

Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,

Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresner, D., Farfan, D.,

Ferrieres, S., Frishe, E., Galle, R.F., Garg, N.S., George, R.A.,

Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,

Ibegwam, C., Jalili, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,

McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,

Pacheco, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,

Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,

Stapleton, M., Strong, R., Sylvestre, R., Tector, C., Williams, S.M.,

Zaveri, J.S., Smith, H.O., Rubin, G.M., and Venter, J.C.

Sequencing of Drosophila chromosome 3R, region 85D-85D

Unpublished

2 (bases 1 to 148847)

Celniker, S.E., Adyan, A., Arcalini, T.T., Baxter, E., Blazek, R.G.,

Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,

Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

DEFINITION	Drosophila melanogaster genomic scaffold 142000013386035 section 7 of 105, complete sequence.
ACCESSION	AE003682 AE002708
KEYWORDS	AE003682.2 GI:10726402
SOURCE	HTG.
ORGANISM	fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 224400)
AUTHORS	Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D., Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Helt,G., Nelson,C.R., Gabor Miklos,G.L., April,J.F., Agbayani,A., An,H.J., Anderson,K., Binkley,C., Baldwin,D., Ballew,R.M., Basu,A., Bendale,S., Bernick,K., Bhatnagar,S., Binkley,R.M., Beeson,K.Y., Benos,P.V., Bereman,B.P., Bhandari,D., Bolshakov,S., Borkov,D., Botchan,M.R., Bouck,J., Brockstein,P., Brothier,P., Burris,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrier,S., Fleischmann,M., Foster,C., Gabriellian,A.E., Gard,N.S., Gehlert,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lascko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Maitel,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Mosniger,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusser,D.R., Pacle,J.M., Palazzolo,M., Peltman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C. The genome sequence of Drosophila melanogaster Science 287 (5461), 2185-2195 (2000)
JOURNAL	2 (bases 1 to 224400)
MEDLINE	Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C. Direct Submission Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA On Oct 9, 2000 this sequence version replaced gi:7299142.
COMMENT	Location/Qualifiers
FEATURES	1..224400 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="3R" complement(join(235..1324,1643..2221,2355..2417)) /gene="dl" /product="CT27541" /db_xref="FLYBASE:FBan0009745" /db_xref="FLYBASE:FBgn0000412" complement(<235..>2417) /gene="dl" /note="CG9745" /map="85D2-85D3" /db_xref="FLYBASE:FBan0009745"
gene	
mrna	
CDS	

JOURNAL	J Bacteriol.	169, 1423-1432 (1967)
MEDLINE	87165745	
FEATURES	source	Location/Qualifiers
		1. .1700
		/organism="Sinorhizobium meliloti"
		/db_xref="taxon:382"
		<1. .255
		/note="nrtrb"
		/codon_start=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRWIAAGDILVTVVMPDENAFDLPILRIKARPDLPYLVAQAQNTFMTAIKASEKGYDYLKPRDLELIGITIGIRALAEPRKRSKLEDSQDQMPVGRSAAMQELRYVARLMTQDILFTMTGESGQKELVARALHDYGRKRRNGPRVPAANMAAPRDLIESLFGKCAFTFGAQRSTRFEQABEGTFLFDEIGDMPDAQTRILRYLQGGFTYVGGRTPRSRVRIYAATKDQSLINQGLFREDYLRNLVPLRPLRDRAEDIPDLVRHVOAEAGGLDVKRDLQALIELMKHPWGPVQVRELENVRLTALYPODVITRELELREIPDSPTIEKAARSGSLISQVAEENMKROYFASGDALPFGLYDRVLAEMEYPLIALATATKGNQIRKAALGLINRTLRKKIRIGLVSYVRSIA"
		252. .1700
		/translation="EFCVHDNGEVPDPDLPLHFDPIITTTKNGSGGLALVAKIIGGGIVIECDQSQHSRTTFRVLPMPASKGLADEETPMTKTNG"
		252. .1700
		/note="nrtrc"
		/codon_start=1
		/transl_table=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRWIAAGDILVTVVMPDENAFDLPILRIKARPDLPYLVAQAQNTFMTAIKASEKGYDYLKPRDLELIGITIGIRALAEPRKRSKLEDSQDQMPVGRSAAMQELRYVARLMTQDILFTMTGESGQKELVARALHDYGRKRRNGPRVPAANMAAPRDLIESLFGKCAFTFGAQRSTRFEQABEGTFLFDEIGDMPDAQTRILRYLQGGFTYVGGRTPRSRVRIYAATKDQSLINQGLFREDYLRNLVPLRPLRDRAEDIPDLVRHVOAEAGGLDVKRDLQALIELMKHPWGPVQVRELENVRLTALYPODVITRELELREIPDSPTIEKAARSGSLISQVAEENMKROYFASGDALPFGLYDRVLAEMEYPLIALATATKGNQIRKAALGLINRTLRKKIRIGLVSYVRSIA"
		252. .1700
		/translation="EFCVHDNGEVPDPDLPLHFDPIITTTKNGSGGLALVAKIIGGGIVIECDQSQHSRTTFRVLPMPASKGLADEETPMTKTNG"
		252. .1700
		/note="nrtrc"
		/codon_start=1
		/transl_table=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRWIAAGDILVTVVMPDENAFDLPILRIKARPDLPYLVAQAQNTFMTAIKASEKGYDYLKPRDLELIGITIGIRALAEPRKRSKLEDSQDQMPVGRSAAMQELRYVARLMTQDILFTMTGESGQKELVARALHDYGRKRRNGPRVPAANMAAPRDLIESLFGKCAFTFGAQRSTRFEQABEGTFLFDEIGDMPDAQTRILRYLQGGFTYVGGRTPRSRVRIYAATKDQSLINQGLFREDYLRNLVPLRPLRDRAEDIPDLVRHVOAEAGGLDVKRDLQALIELMKHPWGPVQVRELENVRLTALYPODVITRELELREIPDSPTIEKAARSGSLISQVAEENMKROYFASGDALPFGLYDRVLAEMEYPLIALATATKGNQIRKAALGLINRTLRKKIRIGLVSYVRSIA"
		252. .1700
		/translation="EFCVHDNGEVPDPDLPLHFDPIITTTKNGSGGLALVAKIIGGGIVIECDQSQHSRTTFRVLPMPASKGLADEETPMTKTNG"
		252. .1700
		/note="nrtrc"
		/codon_start=1
		/transl_table=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRWIAAGDILVTVVMPDENAFDLPILRIKARPDLPYLVAQAQNTFMTAIKASEKGYDYLKPRDLELIGITIGIRALAEPRKRSKLEDSQDQMPVGRSAAMQELRYVARLMTQDILFTMTGESGQKELVARALHDYGRKRRNGPRVPAANMAAPRDLIESLFGKCAFTFGAQRSTRFEQABEGTFLFDEIGDMPDAQTRILRYLQGGFTYVGGRTPRSRVRIYAATKDQSLINQGLFREDYLRNLVPLRPLRDRAEDIPDLVRHVOAEAGGLDVKRDLQALIELMKHPWGPVQVRELENVRLTALYPODVITRELELREIPDSPTIEKAARSGSLISQVAEENMKROYFASGDALPFGLYDRVLAEMEYPLIALATATKGNQIRKAALGLINRTLRKKIRIGLVSYVRSIA"
		252. .1700
		/translation="EFCVHDNGEVPDPDLPLHFDPIITTTKNGSGGLALVAKIIGGGIVIECDQSQHSRTTFRVLPMPASKGLADEETPMTKTNG"
		252. .1700
		/note="nrtrc"
		/codon_start=1
		/transl_table=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRWIAAGDILVTVVMPDENAFDLPILRIKARPDLPYLVAQAQNTFMTAIKASEKGYDYLKPRDLELIGITIGIRALAEPRKRSKLEDSQDQMPVGRSAAMQELRYVARLMTQDILFTMTGESGQKELVARALHDYGRKRRNGPRVPAANMAAPRDLIESLFGKCAFTFGAQRSTRFEQABEGTFLFDEIGDMPDAQTRILRYLQGGFTYVGGRTPRSRVRIYAATKDQSLINQGLFREDYLRNLVPLRPLRDRAEDIPDLVRHVOAEAGGLDVKRDLQALIELMKHPWGPVQVRELENVRLTALYPODVITRELELREIPDSPTIEKAARSGSLISQVAEENMKROYFASGDALPFGLYDRVLAEMEYPLIALATATKGNQIRKAALGLINRTLRKKIRIGLVSYVRSIA"
		252. .1700
		/translation="EFCVHDNGEVPDPDLPLHFDPIITTTKNGSGGLALVAKIIGGGIVIECDQSQHSRTTFRVLPMPASKGLADEETPMTKTNG"
		252. .1700
		/note="nrtrc"
		/codon_start=1
		/transl_table=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRWIAAGDILVTVVMPDENAFDLPILRIKARPDLPYLVAQAQNTFMTAIKASEKGYDYLKPRDLELIGITIGIRALAEPRKRSKLEDSQDQMPVGRSAAMQELRYVARLMTQDILFTMTGESGQKELVARALHDYGRKRRNGPRVPAANMAAPRDLIESLFGKCAFTFGAQRSTRFEQABEGTFLFDEIGDMPDAQTRILRYLQGGFTYVGGRTPRSRVRIYAATKDQSLINQGLFREDYLRNLVPLRPLRDRAEDIPDLVRHVOAEAGGLDVKRDLQALIELMKHPWGPVQVRELENVRLTALYPODVITRELELREIPDSPTIEKAARSGSLISQVAEENMKROYFASGDALPFGLYDRVLAEMEYPLIALATATKGNQIRKAALGLINRTLRKKIRIGLVSYVRSIA"
		252. .1700
		/translation="EFCVHDNGEVPDPDLPLHFDPIITTTKNGSGGLALVAKIIGGGIVIECDQSQHSRTTFRVLPMPASKGLADEETPMTKTNG"
		252. .1700
		/note="nrtrc"
		/codon_start=1
		/transl_table=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRWIAAGDILVTVVMPDENAFDLPILRIKARPDLPYLVAQAQNTFMTAIKASEKGYDYLKPRDLELIGITIGIRALAEPRKRSKLEDSQDQMPVGRSAAMQELRYVARLMTQDILFTMTGESGQKELVARALHDYGRKRRNGPRVPAANMAAPRDLIESLFGKCAFTFGAQRSTRFEQABEGTFLFDEIGDMPDAQTRILRYLQGGFTYVGGRTPRSRVRIYAATKDQSLINQGLFREDYLRNLVPLRPLRDRAEDIPDLVRHVOAEAGGLDVKRDLQALIELMKHPWGPVQVRELENVRLTALYPODVITRELELREIPDSPTIEKAARSGSLISQVAEENMKROYFASGDALPFGLYDRVLAEMEYPLIALATATKGNQIRKAALGLINRTLRKKIRIGLVSYVRSIA"
		252. .1700
		/translation="EFCVHDNGEVPDPDLPLHFDPIITTTKNGSGGLALVAKIIGGGIVIECDQSQHSRTTFRVLPMPASKGLADEETPMTKTNG"
		252. .1700
		/note="nrtrc"
		/codon_start=1
		/transl_table=1
		/protein_id="AAA2634.1"
		/db_xref="GI:152393"
		/translation="MTGATILVADDDAAIRTVLNOALSRAQYDVYRITSMAATLWRW

FEATURES	source
gene	Direct Submission
gene	Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELIIO
gene	EU Consortium
gene	MELIIO EU Consortium
gene	Laboratoire de Biologie Moléculaire des Relations
gene	Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
gene	France, Laboratoire de Génétique et Développement UMR6061-CNRS,
gene	Faculté de Médecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
gene	France, GATC GmbH, Fritz-Arnold-Str. 23, D-78467 Konstanz, Germany,
gene	Université de Bielefeld, Biologie IV (Genetik) Universitätsstr 25,
gene	D-33615 Bielefeld, Germany, Unité de Biochimie physiologique,
gene	Université catholique de Louvain, Place Croix du Sud 2, Bte 20,
gene	B-1348 Louvain-la-Neuve, Belgium, Unité de Microbiologie, Faculté
gene	des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
gene	B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
gene	http://sequense.toulouse.inra.fr/melliocli.html.
gene	Location/Qualifiers
gene	1..329100
gene	/organism="Storhizobium melliocli"
gene	/strain="1021"
gene	/db_xref="taxon:382"
gene	152..227
gene	/gene="tRNA-THR_TGT"
gene	/note="codon recognized: UGU; predicted by tRNAscan-SF"
gene	/product="tRNA-Thr"
gene	/evidence="not_experimental
gene	152..227
gene	/gene="tRNA-THR_TGT"
gene	complement(373..798)
gene	/gene="SMC04435"
gene	complement(373..798)
gene	/gene="SMC04435"
gene	/function="miscellaneous; hypothetical/global homology"
gene	/note="Product confidence : hypothetical
gene	Gene name confidence : hypothetical
gene	predicted by Homology"
gene	/codon_start=1
gene	/transl_table=1
gene	/evidence="not_experimental
gene	/product="CONSERVED HYPOTHETICAL PROTEIN"
gene	/protein_id="CAC45913.1"
gene	/db_xref="GI:15074267"
gene	/translation="MNNALVPEEAVSWMQSKPEYCNILGSCDYERPEEGCYLSLD
gene	GAELMIIDIGRTFADGHLPRSPYFGRLNVIQIRVASDALVRLDRGIALYLPPEE
gene	ERWYRGDEEYGNRFVADPDGYLRFREYPLGRPRTS"
gene	complement(914..1855)
gene	/gene="SMC01330"
gene	complement(914..1855)
gene	/gene="SMC01330"
gene	/function="miscellaneous; not classified regulator"
gene	/note="Product confidence : putative
gene	Gene name confidence : hypothetical
gene	predicted by Codon-usage
gene	predicted by Homology
gene	predicted by Framed"
gene	/codon_start=1
gene	/transl_table=1
gene	/evidence="not_experimental
gene	/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
gene	/protein_id="CAC45914.1"
gene	/db_xref="GI:15074268"
gene	/translation="MODLNDIALFAAAVVKHNGFSAARDLNVPSKLSHYARLEEDD
gene	GVRLERSTRKLRMTTEGRIPEYEAAGVVAAEIAVRAPEYGLACPIGEF
gene	TPMLADLPAHRRYPCVRLITLITASNRIIDLERIDVALAPADLPDLSLYRKGG
gene	EVQRILASPTLLRLRLGETITDNLQMPRTLSMNGCHPNDVRLYHATGALIEHRRY
gene	VGCDFILIERAAIEGMIALLPDHICERARVATALVPVLEWTGVNVMHVLFPSSRR
gene	GLPATRALIDFLAENLIKALERCREVDPRAAFET"
gene	2063..2683
gene	/gene="acpD OR SMC01329"
gene	2063..2683
gene	/gene="acpD OR SMC01329"
gene	/function="small molecule metabolism; fatty acid
gene	biosynthesis"

BASE COUNT	1408 a	821 c	1141 g	944 t
ORIGIN				

Query Match	1.88	Score 39.4	DB 8	Length 3060
Similarity	46.58	Pred. No. 38		
Best Local	0	Mismatches 146	Indels 0	Gaps 0
Matches 127	Conservative			

QY	1811	1239
CTAAGAGCGATGAGTTGGCGGTACGAGAGATATCCTCGACTCTTCGCGACCAAAACT	CTGCGGCGCCATGTTATAGAGTCTCGGAGTGCAGCCCGCAACATATGCAACTCGTCAACT	
1870	1298	

QY	1871	ATGTCGCTATGCCCAGAGCGTCACTCCATATCGCTTTCTTCTTATGGAACAACGACATCA	1930
Db	1299	TGGAGCGCGAATACCGAAACGACAACTCTCCACACACCGACGAGCATCAAGTATATCT	1358

Db 1359 TCCGTCGATTGCGAGGGCAGGACGGCCCTAATGTGAAGAGCAGTGTCTCGAGCTCGGATT 1418

QY	2051	CAATCAGCGTGACATCCGCTACGTTCTCTCA	2083
b	1479	AAGTGGCGGACAAATTTTGGCCATATCTTCA	1511

RESULT 14				
MLEPRTN4	348450 bp	DNA	linear	BCI 20-FEB-2001
LOCUS				
DEFINITION	Mycobacterium leprae strain TN complete genome; segment 4/10.			
ACCESSION	AL583920	AL450380		
VERSION	AL583920.1	GI:13092922		

ORGANISM
Mycobacterium lepre
Bacteria, Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterinae; Mycobacteriaceae;
Mycobacterium.
1 (bases 1 to 348450)
Cole, S.T., Eigmeier, K., Parkhill, J., James, K.D., Thomson, N.R.,
Wheeler, P.R., Honore, N., Ganer, T., Churcher, C., Harris, D.,
Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R.,
Davies, R.M., Dellin, K., Duthoy, S., Felwell, T., Fraser, A.,
Hamlin, N., Holroyd, S., Hornsby, T., Jags, K., Lacroix, C.,

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source

Hamlin, N., Holroyd, S., Hornsby, T., Jagals, K., Lacroix, C.,
 Maclean, J., Moule, S., Murphy, L., Oliver, Quall, M. A.,
 Rajadiream, M. A., Rutherford, K. M., Rutter, S., Seeger, K., Simon, S.,
 Simmonds, M., Skellon, J., Squires, R., Squires, S., Stevens, K.,
 Taylor, K., Whitehead, S., Woodward, J. R. and Barrell, B. G.
 Massive gene decay in the leprosy bacillus
 Nature 409 (6823), 1007-1011 (2001)
 21128732
 2 (bases 1 to 348450)
 Parkhill, J.
 Direct Submission
 Submitted (20-FEB-2001) Submitted on behalf of the Mycobacterium
 leprae sequencing teams, The Sanger Centre, Wellcome Trust Genome
 Campus, Hinxton, Cambridge, CB10 1SA, UK Unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724, Paris Cedex, France. E-mail: parkhill@sanger.ac.uk
 Notes:
 Details of M. leprae sequencing at the Sanger Centre are available
 from http://www.sanger.ac.uk/Projects/M_Leprae/ A relational
 database containing the M. leprae sequences is available from
<http://genolist.pasteur.fr/leproma/>.
 Location/Qualifiers
 I. 348450
 /organism="Mycobacterium leprae"

gene /strain="TN"
/db_xref="taxon:1769"
551..1786
/gene="ML0842"
551..1786
/gene="ML0842"
/note="Similar to Mycobacterium tuberculosis hypothetical
44.6 kDa protein Rv1464 or MT0007.11 TR:053155
(EMBL:AL021184) (417 aa) fasta scores: E(): 0, 43.6% id in
408 aa and to many other bacterial nifs-homologues, e.g.
Bacillus subtilis Yurw protein yurw TR:032164
(EMBL:Z99120) (406 aa) fasta scores: E(): 0, 46.7% id in
405 aa. Previously sequenced as TR:032975 (EMBL:Z98741).
Contains Pfam match to entry PF00266 aminotran_5,
Aminotransferases class-V.
Similar to ML0117, ML0596 and ML1708"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC31223.1"
/db_xref="GI:13092923"
/gb_xref="GI:13092923"
/translation="MVDVFDIQAVRADPFILOFTVNGKPLWEDNAAATYQKPOVIDR
LSFYAHNSNTHRAHMLAARATDAYEARERARPFIGAARAKOELIFRGTEAINL
VYAWGSKHLOGDEVVITLHEHNANIPMOLISQSGAILKVAAPYDDAGNLMSEFE
DLGPRTKLVAATOVSNALGTGTQGEKIVELGHRGAVLIDGAOSIPLPLINVSLE
ADFFVSGHKITGPGIGVYGCEDVITEMPGGGMNIVDTLERSLYCGPKNFE
ACTGNADAVGAGLALRYVERGVORIASHEQALDIYATPRLDIPGVRLNGATKKA
SVLSFVLAGEHPLEVGKALNAGIAVRAGHCAQPVLRRLGLEATVRPSFAFYNTKE
IDVFIVNVRRIAEGGTNG"
653..1738
/gene="ML0842"
/note="Pfam match to entry PF00266 aminotran_5,
Aminotransferases class-V, score 23.70, E-value 1.7e-08"
1308..2395
/gene="lppp"
/note="ML0843"
1908..2395
/gene="ML0843"
/note="Possible pseudogene of M. tuberculosis orthologue
lppp (best blastx score 294)"
/codon_start=1
/pseudo
/transl_table=11
/product="lppoprotein (pseudogene)"
2549..2555
/note="possible RBS"
2559..4112
/gene="nark"
/note="ML0844"
2559..4112
/gene="nark"
/note="Similar to Mycobacterium tuberculosis putative
nitrite extrusion protein nark1 or Rv2329C or MTCY3G12.05
TR:P1883 (EMBL:Z79702) (515 aa) fasta scores: E(): 0,
69.3% id in 488 aa and to Escherichia coli nitrite
extrusion protein 2 nau SW:NAKU_ECOLI (P37758; P77666)
(462 aa) fasta scores: E(): 0, 36.6% id in 459 aa. There
is a frameshift near the C-terminus relative to the M.
tuberculosis homologue. Previously sequenced as TR:032974
(EMBL:Z98741). Contains hydrophobic, probable
membrane-spanning regions."
/codon_start=1
/transl_table=11
/product="putative nitrite extrusion protein"
/protein_id="CAC31225.1"
/db_xref="GI:13092924"
/translation="MDQVLAQAEISLYPGRRSKTVPIDHRHSLSKSFRRLLRRAC
RISHMDPEQAAEAGNKTILRNILMSVTLIGYSVMTIAPVMELFMPKDYGSSA
GPKELIATATLIGACTLRVPSIATATLRGRNMAITSVWLLPTIATWILLAHPLP
LMPYIACALITIGGSGNRPASSTNNAATFPHRLKAGFAGAGNLGVSVIOVMDL
VASVGDRPYIVCGLYVLLITAGIGAMFMNDIEHRIQVNTIRPVLVAVVSTDS
WVIALIYASFSGISFSAFGOVLETFNVAQOSTAQOALHAAELAFIGPTIAAVAR
FWGRLADRLGSGRYTLVVGAMVFAAGLIGYLGITIEGSRVCPIRGVMAVSYPAGFIT

gene complement(4100..4681)
/gene="ML0845"
complement(4100..4681)
/gene="ML0845"
/note="Unknown function. Similar to part of some acyl-CoA
oxidases e.g. Arabidopsis thaliana acyl-CoA oxidase cx2
TR:065201 (EMBL:AF057043) (692 aa) fasta scores: E():
1.5e-06, 35.2% id in 125 aa. Previously sequenced as
TR:032973 (EMBL:Z98741) (193 aa) fasta scores: E(): 0,
29.5% id in 193 aa."
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC31226.1"
/db_xref="GI:13092925"
/translation="MPLRPTNASESRVTRKRTATETLLQWVGARODESKGSLEFNGT
OGRIFFKEDCDILSAVARROYKSEKMFDPFNTYODHLHNAHHTDVRVLEFLA
GISCDPARKLIDIVCDIYALSVARHCTSSITDPSALKAIKAGSRPMPSAAL
CRDVGRRFSESQSSCVTPCRITSRICESELPp"
complement(4959..5480)
/gene="ML0846"
complement(4959..5480)
/gene="ML0846"
/note="Possible pseudogene similar to M. tuberculosis
paralogue Rv1747 (best blastx score 127)"
/codon_start=1
/pseudo
/transl_table=11
/product="ABC transporter (pseudogene)"
complement(5522..5848)
/gene="ML0847"
complement(5522..5848)
/gene="ML0847"
/note="Possible pseudogene of M. tuberculosis orthologue
Rv2327 (best blastx score 184)"
/codon_start=1
/pseudo
/transl_table=11
/product="conserved hypothetical protein (pseudogene)"
5906..8080
/gene="ML0848"
5906..8080
/gene="ML0848"
/note="Similar to Mycobacterium tuberculosis hypothetical
ABC transporter ATP-binding protein Rv2326c or MTCY3G12.08
SW:YN26_MYCTU (P71886) (697 aa) fasta scores: E(): 0,
76.9% id in 697 aa. Shares similar domains with many
ABC-type transporters e.g. Streptomyces roseofulvus ATPase
component of putative ABC transporter frnd TR:068910
(EMBL:AF058302) (524 aa) fasta scores: E(): 8.4e-14, 31.5%
id in 495 aa and Synechococcus sp. nitrate transport
ATP-binding protein NrtD nrtD SW:NRPD_SYNP7 (P38046) (274
aa) fasta scores: E(): 1.6e-12, 34.8% id in 198 aa.
Previously sequenced as TR:032971 (EMBL:Z98741). Contains
hydrophobic, possible membrane-spanning regions. Contains
2 pfam matches to entry PF00005 ABC_tran, ABC transporter.
Contains 2 x PS00017 ATP/GTP-binding site motif A
(P-loop). Contains 2 x PS00211 ABC transporters family
signature."
/codon_start=1
/transl_table=11
/product="ABC transporter"
/protein_id="CAC31229.1"
/db_xref="GI:13092926"
/translation="MDICTHRWALLCOATRCHRWKKTOLGRCAAPDLAPGTGTGKY
TADPRHLHRRSSLOPGLAQASVLAALCATYATYVYPPRAGCALIGYPMGLLA
YTRFREVLTNAAAGVIAFLITGLDGFIAVNSAVIGLIGVYKRRGGTLLVIALA
LFGAGLGAANVAALVLRHLIRKANTAVDGIAAATLTWHLPLWVAVOLKRYFAD
GLQHPWMLIGGYVITILVVSILGMVNLRSVERINDIDVRLDAPSCNEDAPYGP
VPMWLDKVRFRYPHAGDALREVSIDLVRGEHVAVVGANSQGTTLMLITAGREPTSG

[illegible]

	REMARK	Mikrobiologie, Universitaet Dortmund, Fachbereich Chemie/technik,
	REFERENCE	Lehrstuhl fuer Technische Mikrobiologie, Emil-Figge-Str. 66, 44221
	AUTHORS	Dortmund, FRG
JOURNAL	TITLE	revised by [3]
		4 (sites)
	REMARK	Direct Submission
	REFERENCE	Submitted (17-SEP-1996) C. Modara, Institut fuer Technische
	AUTHORS	Mikrobiologie, Universitaet Dortmund, Fachbereich Chemie/technik,
JOURNAL	TITLE	Lehrstuhl fuer Technische Mikrobiologie, Emil-Figge-Str. 66, 44221
		Dortmund, FRG
	REMARK	revised by [4]
	REFERENCE	5 (sites)
	AUTHORS	Friedrich.C.
JOURNAL	TITLE	Direct Submission
		Submitted (25-JUN-1997) Friedrich C., Institut fuer Technische
	REMARK	Mikrobiologie, Universitaet Dortmund, Fachbereich Chemie/technik,
	REFERENCE	Lehrstuhl fuer Technische Mikrobiologie, Emil-Figge-Str. 66, 44221
	AUTHORS	Dortmund, FRG
JOURNAL	TITLE	revised by [5]
		6 (sites)
	REMARK	Friedrich.C.
	REFERENCE	Direct Submission
	AUTHORS	Submitted (10-JUL-1997) Friedrich C., Institut fuer Technische
JOURNAL	TITLE	Mikrobiologie, Universitaet Dortmund, Fachbereich Chemie/technik,
		Lehrstuhl fuer Technische Mikrobiologie, Emil-Figge-Str. 66, 44221
	REMARK	Dortmund, FRG
	REFERENCE	revised by [7]
	AUTHORS	7 (bases 1 to 10624)
JOURNAL	TITLE	Friedrich.C.
		Direct Submission
	REMARK	Submitted (09-JUN-2000) Friedrich C., Institut fuer Technische
	REFERENCE	Mikrobiologie, Universitaet Dortmund, Fachbereich Chemie/technik,
	AUTHORS	Lehrstuhl fuer Technische Mikrobiologie, Emil-Figge-Str. 66, 44221
JOURNAL	TITLE	Dortmund, FRG
		On Jan 24, 2002 this sequence version replaced gi:8517639.
	COMMENT	Location/Qualifiers
FEATURES	SOURCE	1..10624
		/organism="Paracoccus denitrificans"
		/strain="GB17"
		/db_xref="taxon:266"
CDS		complement(320..769)
		/note="ORF1"
		/codon_start=1
		/transl_table=1
		/product="hypothetical protein"
		/protein_id="CAB94376.1"
		/db_xref="GI:8517640"
		/db_xref="SPTRMBL_09LCV3"
		/translation="MIPAPWIRDKLAQSSGCCRAPIYPAFRKKGNDPLGIEINR
		OEPDEQEDPALIKAAANDASTFLKAUGHDRILMLICYLMGPKSVTELEMLSSROY
		VSOOLARLRHREGIVASRRADQTIFYSLIDRPVDLLAVLKLFETDC"
		1228..1965
CDS		/note="ORF2"
		/codon_start=1
		/transl_table=1
		/product="hypothetical protein"
		/protein_id="CAB94377.1"
		/db_xref="GI:8517641"
		/db_xref="SPTRMBL_09LCV2"
		/translation="MDVTYGGAAIAGLSFLSPCLIPMPFLISTWAGSYRELBDG
		GTVFAAGARQMWLRALAFALGVTTTIFLMLDLGTALGAQAFQWKSTLSVAAGVLVF
		GLTFGLVARIGTLVREARYOTEAKPSTLIAGVYMGIAFGGWTPCVGPALAAITLMS
		GMSEIMRGGLLVGYGLAMTLPVLAFFAAPLSWVRARAVMGHVEKMGVLLIVF
		GLIIATNSVANAIADMIRMIRNDFMTATLL"
		1979..2539
CDS		/note="ORF3"
		/codon_start=1
		/transl_table=1
		/product="hypothetical protein"
		/protein_id="CAB94378.1"
		/db_xref="GI:8517642"
		/db_xref="SPTRMBL_09LCV1"

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 19:21:20 ; Search time 249 Seconds

(without alignments)

14748.916 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaatgtaataaaag.....aagactgaagtgtatctaa 2139

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802:*

1: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.2	2.0	1147	23	ABLI4089
2	42.2	2.0	3632	23	ABLI4088
3	40.6	1.9	2868	23	ABLI21554
4	38.8	1.8	4013	23	ABLI18653
5	38.6	1.8	3621	22	AAH48797
6	38.6	1.8	4848	22	AAH48798
7	37.8	1.8	10757	23	ABLI21478
8	37.4	1.7	9845	23	ABLI18652
9	37	1.7	786	22	AAc68978

10	37	1.7	1077	22	AAc68976
11	36.8	1.7	810	22	AAH01671
12	36.8	1.7	34980	22	AAE86431
13	36.4	1.7	1090	21	AAc66137
14	36.4	1.7	1292	22	AAI58976
15	36.4	1.7	1649	21	AA235391
16	36.4	1.7	1768	22	AAI60762
17	36.2	1.7	345	22	AAH67046
18	36.2	1.7	536	21	AAH3983
19	36.2	1.7	2280	17	AAI36127
20	36.2	1.7	34980	22	AAH68530
21	35.6	1.7	6174	22	AAK81576
22	35.6	1.7	8617	22	AAK81577
23	35.6	1.7	32313	22	AAK81578
24	35.2	1.6	449	20	AAV68139
25	35.2	1.6	4590	22	AAH24065
26	35	1.6	1910	23	AAH85131
27	34.8	1.6	632	21	AAH08951
28	34.8	1.6	5059	20	AAH84332
29	34.6	1.6	639	21	AAH3795
30	34.6	1.6	975	23	ABLI6623
31	34.6	1.6	2975	23	ABLI6622
32	34.6	1.6	5133	23	ABLI09568
33	34.2	1.6	3059	8	AAH70566
34	34	1.6	2759	23	ABLI06895
35	34	1.6	3972	24	AAH97214
36	34	1.6	5837	12	AAO14461
37	34	1.6	5837	14	AAO43966
38	34	1.6	5837	14	AAO43515
39	34	1.6	5837	15	AAO66293
40	34	1.6	5837	17	AAI17546
41	34	1.6	5837	18	AAI70812
42	34	1.6	5837	20	AAV68141
43	34	1.6	5837	23	AAH60490
44	34	1.6	13620	23	ABLI06894
45	34	1.6	236303	22	AAH11614

ALIGNMENTS

RESULT 1	ABLI4089	ABLI4089 standard; cDNA: 1147 BP.
ID	ABLI4089	
AC	ABLI4089;	
DT	26-MAR-2002 (first entry)	
XX		
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 36749.	
XX		
KW	Drosophila; developmental biology; cell signalling; insecticide;	
KW	Pharmaceutical; gene; ss.	
XX		
OS	Drosophila melanogaster.	
XX		
PN	WO200171042-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	23-MAR-2001; 2001WO-US09231.	
XX		
PR	23-MAR-2000; 2000US-191637P.	
XX		
PR	11-JUL-2000; 2000US-0614150.	
XX		
PA	(PEKE) PE CORP NY.	
XX		
PI	Venter JC, Adams M, Li PWD, Myers EW;	
XX		
DR	WPI; 2001-656860/75.	
XX		
DR	P-PSDB; ABB69986.	
XX		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more	

PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
PS Claim 1; SEQ ID NO 36749; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB101840-AB116175).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1147 BP; 257 A; 335 C; 303 G; 252 T; 0 other;
Query Match 2.0%; Score 42.2; DB 23; Length 1147;
Best Local Similarity 49.3%; Pred. No. 0.085;
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 1470 GGTTCCTTATAGCGACAAGTTCATGCCATGCTCAAGTCATGACAAAGAAAGTTTGC 1529
DB 711 GGCCGCTATACGCCGCTGCTGCTGCCCTCCAGAAATCCAGCGAGTTCCTTGT 770
QY 1530 CAAGGCTATCGAAGAAAGATCCGCGAGTAGAGCTTCCAAAGCGCTAATAGCTGCTCG 1589
DB 771 CCAACACCGAGGAAGACGCGAGTAGAGCTGCGAGCGATGATACATTCGCGCTTC 830
QY 1590 CGCTATTACGAGCCGATGCGCAATGCTATGCCATGAGAAAGGCAAGCGCTTTT 1649
DB 831 CACCAACCAATCCCTACCAACCTGCGGCTCTACTAGGACGACACCGGATCTTCAA 890
QY 1650 CTTTGGCGGTTTGGCTGAGATGTACCCCGAGCGTCTGCGC 1692
DB 891 CATCATCCTGTGTTCATGTCGCTTTCGACTGTCTCTGCTG 933
RESULT 2
AB114088
ID ABL14088 standard; cDNA; 3632 BP.
XX
AC ABL14088;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36746.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT P-PSDB; ABB69985.
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -
XX
PS Claim 1; SEQ ID NO 36746; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB101840-AB116175), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB101840-AB116175).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3632 BP; 1062 A; 799 C; 773 G; 998 T; 0 other;
Query Match 2.0%; Score 42.2; DB 23; Length 3632;
Best Local Similarity 49.3%; Pred. No. 0.16;
Matches 110; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
QY 1470 GGTTCCTTATAGCGACAAGTTCATGCCATGCTCAAGTCATGACAAAGAAAGTTTGC 1529
DB 2196 GGCCGCTATACGCCGCTGCTGCTGCCCTCCAGAAATCCAGCGAGTTCCTTGT 2255
QY 1530 CAAGGCTATCGAAGAAAGATCCGCGAGTAGAGCTTCCAAAGCGCTAATAGCTGCTCG 1589
DB 2256 CCAAGCAACGGAAGAAAGACGTCGAGCATGCGTGCCAAAGGATATACCATTCGCGCTTC 2315
QY 1590 CGCTATTACGAGCCGATGCGCAATGCTATGCCATGAGAAAGGCAAGCGCTTTT 1649
DB 2316 CACCAACCAATCCCTACCAACCTGCGGCTCTACTAGGACGACACCGGATCTTCAA 2375
QY 1650 CTTTGGCGGTTTGGCTGAGATGTACCCCGAGCGTCTGCGC 1692
DB 2376 CATCATCCTGTGTTCATGTCGCTTTCGACTGTCTCTGCTG 2418
RESULT 3
AB121554
ID ABL21554 standard; DNA; 2868 BP.
XX
AC ABL21554;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 16135.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX


```
FT      /product= "soxa"
XX
XX      DE1006201-A1.
XX
XX      16-AUG-2001.
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      (FRIE/) FRIEDRICH C.
XX
XX      Friedrich C;
XX
XX      WPI: 2001-497622/55.
XX
XX      P-PSDB: AAB86529, AAB86530, AAB86531, AAB86532, AAB86533, AAB86534.
XX
XX
XX      New genes involved in oxidation of reduced sulphur compounds,
XX      particularly conversion of thiosulphate to sulphate, and related
XX      proteins, from Paracoccus pantotrophus -
XX
XX      Claim 1; Page 10-14; 22pp: German.
XX
XX      This invention describes novel genes (I) of the sox (sulphur oxidation)
XX      region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA
XX      from Paracoccus pantotrophus GB17. The proteins (II) encoded by (I) are
XX      used for oxidation of reduced sulphur compounds in cellular or cell-free
XX      systems, particularly of thiosulphate to sulphate. (I) is also used for
XX      expression of the proteins and to prepare transgenic plants or animals
XX      or transformed microorganisms. This sequence encodes the ORF2, ORF3 and
XX      the soxXYZA proteins from the P. pantotrophus sox region.
XX
XX      Sequence 3621 BP; 650 A; 1153 C; 1186 G; 632 T; 0 other;
SQ
Query Match      1.8%; Score 38.6; DB 22; Length 3621;
Best Local Similarity 48.8%; Pred. No. 1.8;
Matches 104; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
OY      1847 TCGACCTCTTCGCGACCAAAACATATGGTGCCTATGCCGAGACGGTGCATTCATTCG 1906
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      3026 TCGAGCGCGCATCAGCAAGACATGCGGGGTTGTGGAATGCTCCCGCTCAAGATCC 2967
OY      1907 CTTTCCTATCGAACACGACATCAGCGGGGTAACTCCGGTAGCCCCGATTTCGATAAGA 1966
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      2966 CGCGTGTGTGCTGCGGAAGTGCCAGCCGGAATGATGTGCAACGATACGGCCAGA 2907
OY      1967 ACGGCGGTGTGATGGTCTGCTTTCGATGGCACTGGGAAGCTATGAGTGTGATCAG 2026
        ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB      2906 AAGGCGGGGCGGTGCGTCTTGGTCAAGATCTTCACCGGTCGCTGCGTTTCGATGACC 2847
OY      2027 AGTTGAACCCGATCTGCAGCGACACATCAGCG 2059
        || | | | | | | | | | | | | | | | | | | | | |
DB      2846 AGCCGTCCTCGACCGGTCGCCGAAGGCCGCG 2814

RESULT 6
AAH48798/C
ID      AAH48798 standard; DNA: 4848 BP.
XX
XX      AAH48798;
XX
XX      05-NOV-2001 (first entry)
XX
XX      P. pantotrophus GB17 DNA encoding sox-associated proteins.
XX
XX      Sulphur oxidation; sox; ORF1; ORF2; ORF3; soxX; soxY; soxZ; soxA;
XX      reduced sulphur; thiosulphate; transgenic plant; transgenic animal; ds.
XX
XX      Paracoccus pantotrophus.
XX
XX      DE1006201-A1.
XX
XX      16-AUG-2001.
XX
```

```
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      11-FEB-2000; 2000DE-1006201.
XX
XX      (FRIE/) FRIEDRICH C.
XX
XX      Friedrich C;
XX
XX      WPI: 2001-497622/55.
XX
XX
XX      New genes involved in oxidation of reduced sulphur compounds,
XX      particularly conversion of thiosulphate to sulphate, and related
XX      proteins, from Paracoccus pantotrophus -
XX
XX      Claim 1; Fig 1; 22pp: German.
XX
XX      This invention describes novel genes (I) of the sox (sulphur oxidation)
XX      region designated ORF1, ORF2, ORF3 (Open Reading Frame), and soxXYZA
XX      from Paracoccus pantotrophus GB17. The proteins (II) encoded by (I) are
XX      used for oxidation of reduced sulphur compounds in cellular or cell-free
XX      systems, particularly of thiosulphate to sulphate. (I) is also used for
XX      expression of the proteins and to prepare transgenic plants or animals
XX      or transformed microorganisms. This sequence encodes the P. pantotrophus
XX      sox region associated proteins described in the invention.
XX
XX      Sequence 4848 BP; 885 A; 1584 C; 1552 G; 827 T; 0 other;
SQ
Query Match      1.8%; Score 38.6; DB 22; Length 4848;
Best Local Similarity 48.8%; Pred. No. 2.1;
Matches 104; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
OY      1847 TCGACCTCTTCGCGACCAAAACATATGGTGCCTATGCCGAGACGGTGCATTCATTCG 1906
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      4253 TCGAGCGCGCATCAGCAAGACATGCGGGGTTGTGGAATGCTCCCGCTCAAGATCC 4194
OY      1907 CTTTCCTATCGAACACGACATCAGCGGGGTAACTCCGGTAGCCCCGATTTCGATAAGA 1966
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      4193 CGCGTGTGTGCTGCGGAAGTGCCAGCCGGAATGATGTGCAACGATACGGCCAGA 4134
OY      1967 ACGGCGGTGTGATGGTCTTTCGATGGCACTGGGAAGCTATGAGTGTGATCAG 2026
        ||| | | | | | | | | | | | | | | | | | | | | | | | | |
DB      4133 AAGGCGGGGCGGTGCGTCTTGGTCAAGATCTTCACCGGTCGCTGCGTTTCGATGACC 4074
OY      2027 AGTTGAACCCGATCTGCAGCGACACATCAGCG 2059
        || | | | | | | | | | | | | | | | | | | | | |
DB      4073 AGCCGTCCTCGACCGGTCGCCGAAGGCCGCG 4041

RESULT 7
ABL21478
ID      ABL21478 standard; DNA: 10757 BP.
XX
XX      ABL21478;
XX
XX      26-MAR-2002 (first entry)
XX
XX      Drosophila melanogaster genomic polynucleotide SEQ ID NO 15907.
XX
XX      Drosophila; developmental biology; cell signalling; insecticide;
XX      pharmaceutical; gene; ds.
XX
XX      Drosophila melanogaster.
XX
XX      WO200171042-A2.
XX
XX      27-SEP-2001.
XX
XX      23-MAR-2001; 2001WO-US09231.
XX
XX      23-MAR-2000; 2000US-191637P.
XX
XX      11-JUL-2000; 2000US-0614150.
XX
```


PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW.
 XX WPI: 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Claim 1; SEQ ID NO 15907; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 10757 BP; 3125 A; 2268 C; 2226 G; 3138 T; 0 other;
 Query Match 1.8%; Score 37.8; DB 23; Length 10757;
 Best Local Similarity 54.1%; Pred. No. 5.4;
 Matches 100; Conservative 0; Mismatches 82; Indels 3; Gaps 1;
 OY 1283 AATGCTTGACGACAGTACAAAGACTACCTCCCTGCTGACGCTAAGGCTGCGCG 1342
 DB 6709 ATTGTTTTCGACATGACCAAGCTCTGATGCCATTTGAGGAAAGGTTCAGCCGG 6768
 OY 1343 CCATGCTGATATTGTACCGCGGTATCCCTGCCAGACCTCCCGATATATTCAGA 1402
 DB 6769 GT---GACTTATCATTCGCGGAGGTGACGAGGATTAATTTATGATGCT 6825
 OY 1403 ATGTAATGACAGAAATTCAAAGCGGACGACGAGATATGACACTTTCGATTCGACA 1462
 DB 6826 AAGTAAACAAACAAATTTCATCAACCTCATGATATATGACACACTTCTGTTTGCA 6885
 OY 1463 AGAGT 1467
 DB 6886 AAGT 6890

RESULT 8
 ABL18652/c
 ID ABL18652 standard; DNA; 9845 BP.
 XX
 AC ABL18652;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7429.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX

PI Venter JC, Adams M, Li PWD, Myers EW.
 XX WPI: 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Claim 1; SEQ ID NO 7429; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 9845 BP; 2751 A; 2130 C; 2204 G; 2760 T; 0 other;
 Query Match 1.7%; Score 37.4; DB 23; Length 9845;
 Best Local Similarity 49.7%; Pred. No. 6.8;
 Matches 95; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
 OY 1294 GACAAATCAAAAGTACTACCTCCCGCTGACACGCTAGTGTGCGCCCATGCTCGAT 1353
 DB 8624 GACATATATAGGAGTGTGCTCATTTGATGTGCTGATATATTCGCGGTGAGATGCC 8565
 OY 1354 ATGTAGCCGCGGTATCCCTGCCGACAACTCCCGATATATTCAGATATATGAC 1413
 DB 8564 ATGGGTGACAGCTGTGTGCTGACGACAAACCGGCGGACCAACGCGATATATCAT 8505
 OY 1414 AAGAATTCAAAGGAGGACGACAGAGATGACAGACTTGTATTCGACAGAGTGTGTT 1473
 DB 8504 CGAATGCTTTCCTGCGCTTTCAGAGATGATCTGAGCTGATTTGACGAGACTTTCGT 8445
 OY 1474 CCTTATACGA 1484
 DB 8444 CCGAGGCGCA 8434

RESULT 9
 AAC68978
 ID AAC68978 standard; DNA; 786 BP.
 XX
 AC AAC68978;
 XX
 DT 27-FEB-2001 (first entry)
 XX
 DE Merozoite surface protein-133 coding sequence.
 XX
 KW Merozoite surface protein; protozoacide; vaccine; malaria; ss.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200063245-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 20-APR-2000; 2000WO-GB01558.
 XX
 PR 20-APR-1999; 99GB-0009072.
 XX
 PR 13-MAY-1999; 99US-0311817.
 XX
 PR 25-MAY-1999; 99CA-2271451.
 XX
 PA (MEDT-) MEDICAL RES COUNCIL.
 XX
 PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthairipbull C;
 XX

DR WPI; 2001-015762/02.
 DR P-PSDB; AAB37610.
 XX Novel variants of the C-terminal fragment of Plasmodium merozoite
 PT surface protein-1, useful as vaccines for treating or preventing
 PT malaria -
 XX
 PS Example 5; Fig 15; 126pp; English.
 XX The present invention relates to non-natural variants of a C-terminal
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 CC non-natural variants have reduced affinity for at least 1 antibody
 CC capable of blocking a second antibody that inhibits the proteolytic
 CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
 CC MSP-1.4.2, compared to natural MSP-1.9. The non-natural variants of the
 CC present invention are useful for immunising a mammal against malaria, and
 CC can be used to treat malaria. The present sequence is the MSP-133
 CC coding sequence.
 XX
 SQ Sequence 786 BP; 245 A; 164 C; 159 G; 218 T; 0 other;
 Query Match 1.7%; Score 37; DB 22; Length 786;
 Best Local Similarity 52.2%; Pred. No. 2.4;
 Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 1377 CGACACAGCTCCCGATATATTCAGAAATGTAATCGACAGAAATTCAAAGCGACACGAA 1436
 DB 135 CGTTAAGCTCAGACATTTTGAACCTCCAGATTCAACAGAGAGAAAACCTCAAGAACGT 194
 QY 1437 GAAGTATGACAGACTTCGTATTCGACAAAGAGTGTGCTTATAGCGACAAAGTTCCATGC 1496
 DB 195 TCTGAGAGCTGACTTGATTCATCAAGAGATTGACTTCTTAACACTGTTTAAGA 254
 QY 1497 CATGCTCAAGTCATGACAGAAAGTTGCCAAG 1533
 DB 255 CCCATACAAAGTTCCTGAACAGAGAGAGACAAAG 291
 RESULT 10
 AAC68976
 ID AAC68976 standard; DNA: 1077 BP.
 XX
 AC AAC68976;
 XX
 DT 27-FEB-2001 (first entry)
 XX
 DE Merozoite surface protein-142 coding sequence.
 XX
 KM Merozoite surface protein; protazoxide; vaccine; malaria; ss.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200063245-A2.
 XX
 PD 26-OCT-2000.
 XX
 PF 20-APR-2000; 2000WO-GB01558.
 XX
 PR 20-APR-1999; 99GB-0009072.
 PR 13-MAY-1999; 99US-0311817.
 PR 25-MAY-1999; 99CA-2271451.
 XX
 PA (MEDT-) MEDICAL RES COUNCIL.
 XX
 PI Holder A, Birdsall B, Feeney J, Morgan W, Syed S, Uthaiyibull C;
 DR WPI; 2001-015762/02.
 XX
 PT Novel variants of the C-terminal fragment of Plasmodium merozoite
 PT surface protein-1, useful as vaccines for treating or preventing
 PT malaria -
 XX

PS Example 5; Fig 15; 126pp; English.
 XX The present invention relates to non-natural variants of a C-terminal
 CC fragment of a Plasmodium merozoite surface protein-1 (MSP-1). The
 CC non-natural variants have reduced affinity for at least 1 antibody
 CC capable of blocking a second antibody that inhibits the proteolytic
 CC cleavage of Plasmodium MSP-1.4.2, and has the same affinity for at least
 CC one third antibody that inhibits the proteolytic cleavage of Plasmodium
 CC MSP-1.4.2, compared to natural MSP-1.9. The non-natural variants of the
 CC present invention are useful for immunising a mammal against malaria, and
 CC can be used to treat malaria. The present sequence is the MSP-142
 CC coding sequence.
 XX
 SQ Sequence 1077 BP; 338 A; 228 C; 225 G; 286 T; 0 other;
 Query Match 1.7%; Score 37; DB 22; Length 1077;
 Best Local Similarity 52.2%; Pred. No. 2.8;
 Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 1377 CGACACAGCTCCCGATATATTCAGAAATGTAATCGACAGAAATTCAAAGCGACACGAA 1436
 DB 135 CGTTAAGCTCAGACATTTTGAACCTCCAGATTCAACAGAGAGAAAACCTCAAGAACGT 194
 QY 1437 GAAGTATGACAGACTTCGTATTCGACAAAGAGTGTGCTTATAGCGACAAAGTTCCATGC 1496
 DB 195 TCTGAGAGCTGACTTGATTCATCAAGAGATTGACTTCTTAACACTGTTTAAGA 254
 QY 1497 CATGCTCAAGTCATGACAGAAAGTTGCCAAG 1533
 DB 255 CCCATACAAAGTTCCTGAACAGAGAGAGACAAAG 291
 RESULT 11
 AAH01671/C
 ID AAH01671 standard; DNA: 810 BP.
 XX
 AC AAH01671;
 XX
 DT 24-JUL-2001 (first entry)
 XX
 DE Fonsecaea pedrosi nucleotide sequence SEQ ID NO:1664.
 XX
 KM Species specific; genus specific; family specific; probe; detection;
 KM identification; algal; archaeal; bacterial; fungal; parasitical;
 KM microorganism; diagnosis; translation elongation factor Tu; toxin;
 KM translation elongation factor G; RecA recombinase; resistance;
 KM catalytic subunit of proton-translocating ATPase; antimicrobial;
 KM vaccine; primer; ds.
 XX
 OS Fonsecaea pedrosol.
 XX
 PN WO200123604-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-CA01150.
 XX
 PR 28-SEP-1999; 99CA-2283458.
 PR 19-MAY-2000; 2000CA-2307010.
 XX
 PA (INFE-) INFECTIO DIAGNOSTIC (IDI) INC.
 XX
 PI Bergeron MG, Boissinot M, Huletsky A, Menard C, Ouellette M;
 PI Picard FU, Roy PH.
 DR WPI; 2001-245006/25.
 XX
 PT Nucleic acid sequences are used to generate universal probes and
 PT primers which can be used to identify and detect the presence of algal,
 PT archaeal, bacterial, fungal and parasitological species in a test sample -
 XX
 XX Claim 24; Page 1264-1265; 1580pp; English.

The present invention describes a method for generating a repository of nucleic acids of tuf, fus, atp and/or recA genes from which probes and/or primers are derived. The method comprises amplifying the nucleic acids of determined algal, archaeal, bacterial, fungal and parasitic species with a combination of defined primer pairs. The method can be used for producing probes and/or primers for detecting one or more related microorganisms e.g. algae, archaea, bacteria, fungi and parasites, for universal detection and for specific and ubiquitous detection and identification of an algal, archaeal, bacterial, fungal and parasitic species, genus, family and group. A nucleic acid (1) obtained using the method of the invention can be used for the universal detection of any bacterium, fungus or parasite in a sample and for the detection of at least one antimicrobial agent resistance gene or at least one toxin gene. hexa nucleic acids are used for the specific and ubiquitous detection and for identification of Streptococcus pneumoniae. (1) can be used to design a therapeutic agent which is effective against microorganisms. Microbial species or genus or family or phylum or group which can be detected include Abiotrophia adiacens, Bordetella sp., Corynebacterium sp., Enterobacteriaceae group, Escherichia coli, Mycobacteriaceae family, Pseudomonads group, Streptococcus sp., Neisseria gonorrhoeae and Staphylococcus sp.. Using DNA based tests provides faster results than substrate specificity tests as results can be determined in an hour and improved accuracy is also achieved. AAH00010 to AAH002304 represent nucleotide sequences and primers/probes which are given in the exemplification of the present invention.

Sequence 810 BP; 167 A; 254 C; 228 G; 161 T; 0 other;

Query Match 1.7%; Score 36.8; DB 22; Length 810;
Best Local Similarity 46.2%; Pred. No. 2.8;
Matches 122; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 192 TGGCGGTACTTCTGGGTGGGAGATGACCGGTATCACAGTGTCCGATCAGGGCCTGAT 251
DB 305 TGGCGGTACTTCTGGGTGGGAGATGACCGGTATCACAGTGTCCGATCAGGGCCTGAT 246
QY 252 CTTTACCAACCACTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 311
DB 245 TCTTTCGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 186
QY 312 TATTCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
DB 185 CGATGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 126
QY 372 TTTCCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
DB 125 CGACGAGTCCAACTCTCTGCTCAATGATGATGATGATGATGATGATGATGATGAT 66
QY 432 GGGTATCACTGAGAGATGAGAG 455
DB 65 TGGCGTTCGACCGACCTGGCGGG 42

RESULT 12
AAF86431/C
ID AAF86431 standard; DNA; 349980 BP.

AC AAF86431;

DT 29-OCT-2001 (first entry)

DE Pyrococcus abyssi genomic fragment #1.

KM Hyperthermophilic archaeon; hyperthermophilic protein; ds.

OS Pyrococcus abyssi.

Key Location/Qualifiers
misc_feature 300001..349980

/tag= a
/note= "This sequence overlaps with the 5' end of
AAH41223"

PN FR2792651-A1.
XX 27-OCT-2000.
PD 21-APR-1999; 99FR-0005034.
XX 21-APR-1999; 99FR-0005034.
PF 21-APR-1999; 99FR-0005034.
PR (CNRS) CNRS CENT NAT RECH SCT.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
PA Forterre P, Thierry JC, Prieur D, Dietrich J, Lecomte O;
PI Querrelou J, Weissenbach J, Saurin W, Hellig R;
XX WPI: 2001-126236/14.
DR New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
PT Claim 1; Page 183-279; 1657pp; French.

XX The present invention relates to the genomic sequence of Pyrococcus
CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC vents. The present sequence is a fragment of the genomic sequence of P.
CC abyssi. The 3' end of this sequence overlaps with the 5' end of AAH41223.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO200065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAB66436.

Sequence 349980 BP; 95167 A; 78444 C; 78870 G; 97499 T; 0 other;

Query Match 1.7%; Score 36.8; DB 22; Length 349980;
Best Local Similarity 49.5%; Pred. No. 65;
Matches 95; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1377 CGACAACTCCCGATATATTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1436
DB 168164 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 168105
QY 1437 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1496
DB 168104 GGAATAAG 168045
QY 1497 CATGCTCAAGTCCATGACAG 1556
DB 168044 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 167985
QY 1557 AGAGCTTTCCAA 1568
DB 167984 CCAGCTCGCCGA 167973

RESULT 13
AAC66137
ID AAC66137 standard; DNA; 1090 BP.

AC AAC66137;

DT 13-FEB-2001 (first entry)

DE Human kidney disease associated gene SEQ ID 7.

KM Nephrotropic; cytosolic; human; kidney disease associated; hypertension;
KM Bartter's syndrome; Gitelman syndrome; nephrolithiasis; renal failure;
KM renal amyloidosis; primary aldosteronism; Addison's disease; cancer;
KM glomerulonephritis; dysplastic malformation; medullary cystic disease;
KM medullary sponge kidney; tubular dysplasia; Alport's syndrome; ds.

OS Homo sapiens.

Db 459 CGTGAACCTGCTGATGGCTGTGATGATGAGCCGAGAGTAGATTTCGCCAATCTC 518
 QY 1854 CTTCCGCACC 1863
 Db 519 CTTCCATGCC 528
 RESULT 15
 AA235391
 ID AA235391 standard; cDNA; 1649 BP.
 AC AA235391;
 XX
 DT 11-APR-2000 (first entry)
 XX
 DE Human kidney-specific protein KK86 cDNA.
 XX
 KW KK86; human; renal failure; kidney disease; glomerulonephritis;
 KM hypertension; polycystic renal disease; antidiabetic; nephrotropic;
 XX diagnosis; gene therapy; vaccine; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 578..1435
 FT /*tag= a
 FT
 XX
 PN MO964435-A1.
 XX
 PD 16-DEC-1999.
 XX
 PF 27-MAY-1999; 99WO-US11757.
 XX
 PR 10-JUN-1998; 98US-0095627.
 XX
 PA (SMIR) SMITHKLINE BEECHAM CORP.
 XX
 PI Chen Z, Fredrickson T, Hu E;
 XX
 DR WPI; 2000-116519/10.
 DR P-PSDB; AAY58646.
 XX
 PT New human kidney specific gene KK86 polypeptides, used to identify
 PT agonists, antagonists and inhibitors for use in therapy -
 XX
 PS Claim 2; Page 27; 33pp; English.
 XX
 CC The present sequence represents cDNA coding for human KK86, a novel
 CC protein whose expression is highly restricted to the kidney, and
 CC which may play a role in kidney-related diseases. The invention
 CC provides polynucleotides having at least 70% identity with the
 CC present sequence, as well as KK86 polypeptides, recombinant
 CC materials and methods for their production. KK86 polypeptides may
 CC be used for identifying agonists and antagonists/inhibitors, and
 CC for detecting diseases associated with inappropriate KK86 activity
 CC or levels. KK86 polypeptides and polynucleotides, agonists,
 CC antagonists and antibodies are used to treat kidney-related
 CC diseases, such as acute or chronic renal failure, diabetes, kidney
 CC disease (including glomerulonephritis), hypertension, and polycystic
 CC renal disease. The polynucleotides are also useful as sources of
 CC primers and probes which can be used to identify mutations in the
 CC KK6 gene, and also for detecting the above diseases.
 XX
 SQ Sequence 1649 BP; 376 A; 474 C; 436 G; 363 T; 0 other;
 Query Match 1.7%; Score 36.4; DB 21; Length 1649;
 Best Local Similarity 49.5%; Pred. No. 5.2;
 Matches 94; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
 QY 1674 CCCGGACGTGCTCTGCCGAGCAGTCCACATCGATGAGCTACGCGCTCCAT 1733
 Db 685 CCGAGACCGTGTCTTACACACCTACAGCTCAGACACGACACAGACAGTGGACTTCGT 744

QY 1734 CAAGGATATGAACCCGAGAGCGGTGCTGTACAACTATCATACAGCAAGGCGCT 1793
 Db 745 CAGAGCAAGCATGCCCGAGTTTGGGGCTTCTCTACAGAAATGACAGTCAAGGAGC 804
 QY 1794 ATTGAGAACCCAGAGATCCTAAGAGCGATGATTTGCCGTACAGAGAAATATCCTCGACT 1853
 Db 805 CGTGGACCTGCTGATGGCGCTGTGATGATGATGAGTGGAGCCGAGAGTAAATTTCCCAACTC 864
 QY 1854 CTTCCGCACC 1863
 Db 865 CTTCCATGCC 874

Search completed: October 17, 2002, 20:33:09
 Job time : 828 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 20:13:20 : Search time 60 seconds
(without alignments) 8756.834 Million cell updates/sec

Title: US-10-008-355-1

Perfect score: 2139

Sequence: 1 atgcaatgaattaaaaag.....aagagctgaattgacctaa 2139

Scoring table:

IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36.4	1.7	1090	US-09-289-349-7	Sequence 7, Appl
2	36.2	1.7	2280	US-08-415-818-5	Sequence 5, Appl
3	36.2	1.7	2280	US-08-894-236-5	Sequence 5, Appl
4	36.2	1.7	2280	PCR-US96-0144-5	Sequence 5, Appl
5	36.2	1.7	7218	US-08-232-463-14	Sequence 14, Appl
6	35.2	1.6	449	US-08-825-782-2	Sequence 2, Appl
7	34	1.6	5837	US-07-686-340-1	Sequence 1, Appl
8	34	1.6	5837	US-08-004-139B-1	Sequence 1, Appl
9	34	1.6	5837	US-08-117-491-1	Sequence 1, Appl
10	34	1.6	5837	US-08-271-364A-1	Sequence 1, Appl
11	34	1.6	5837	US-08-811-492-1	Sequence 1, Appl
12	34	1.6	5837	US-08-222-715B-1	Sequence 1, Appl
13	34	1.6	5837	PCR-US96-10545A-1	Sequence 1, Appl
14	33.6	1.6	50937	US-09-773-816-1	Sequence 1, Appl
15	33.6	1.6	50937	US-09-428-517-1	Sequence 1, Appl
16	33.2	1.6	939	US-09-105-390-45	Sequence 45, Appl
17	33.2	1.6	1020	US-09-105-390-61	Sequence 61, Appl
18	33.2	1.6	2612	US-09-105-390-7	Sequence 7, Appl
19	32.4	1.5	710	US-08-998-416-603	Sequence 603, App
20	32.2	1.5	3468	US-07-951-715A-2	Sequence 2, Appl
21	32.2	1.5	3468	US-08-459-448A-2	Sequence 2, Appl
22	32.2	1.5	3468	US-08-459-595A-3	Sequence 3, Appl
23	32.2	1.5	3468	US-08-459-504B-2	Sequence 2, Appl
24	32.2	1.5	3468	US-08-459-444-2	Sequence 2, Appl
25	32.2	1.5	3468	US-09-053-549-3	Sequence 3, Appl
26	32.2	1.5	3468	US-09-547-422-2	Sequence 2, Appl
27	32	1.5	289	US-09-007-005-17	Sequence 17, Appl

28	32	1.5	289	US-09-244-796-17	Sequence 17, Appl
29	32	1.5	823	US-08-998-416-551	Sequence 551, App
30	32	1.5	1947	US-07-951-715A-3	Sequence 3, Appl
31	32	1.5	1947	US-08-459-448A-3	Sequence 3, Appl
32	32	1.5	1947	US-08-459-595A-3	Sequence 3, Appl
33	32	1.5	1947	US-08-459-504B-3	Sequence 3, Appl
34	32	1.5	1947	US-08-459-444-3	Sequence 3, Appl
35	32	1.5	1947	US-09-547-422-3	Sequence 3, Appl
36	32	1.5	2845	US-08-289-653-1	Sequence 1, Appl
37	32	1.5	3468	US-07-951-715A-4	Sequence 4, Appl
38	32	1.5	3468	US-07-951-715A-8	Sequence 8, Appl
39	32	1.5	3468	US-08-459-448A-4	Sequence 4, Appl
40	32	1.5	3468	US-08-459-595A-4	Sequence 4, Appl
41	32	1.5	3468	US-08-459-504B-4	Sequence 4, Appl
42	32	1.5	3468	US-08-459-504B-8	Sequence 8, Appl
43	32	1.5	3468	US-08-459-504B-4	Sequence 4, Appl
44	32	1.5	3468	US-08-459-504B-8	Sequence 8, Appl
45	32	1.5	3468	US-08-459-444-4	Sequence 4, Appl

ALIGNMENTS

```
RESULT 1
US-09-289-349-7
; Sequence 7, Application US/09289349
; Patent No. 6277574
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael, G.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Klingert, Tod, M.
; APPLICANT: Azimzai, Yalda
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: GENES ASSOCIATED WITH DISEASES OF THE KIDNEY
; FILE REFERENCE: PB-0010 US
; CURRENT APPLICATION NUMBER: US/09/289,349
; CURRENT FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 1090
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 2673009CB1
US-09-289-349-7

Query Match      1.7%: Score 36.4; DB 4; Length 1090;
Best Local Similarity 49.5%; Pred. No. 0.18;
Matches 94; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1674 CCCCGAGCTGCTCTCCGAGGATGCCAATTCACCATGCGTATGAGTACGCTCAT 1733
    || || || || || || || || || || || || || || || || || || || ||
DB 136 CTTGACCGTGTCTTTCACCATGAGCTCATGACGACGACGACGACGAGTGGATTCG 195

QY 1734 CAAAGGATATGAACCCAGAGAGCGTCCCTGTRCACTATCATACAGCAGGAGGCGT 1793
    || || || || || || || || || || || || || || || || || || || ||
DB 196 CAGGACCAAGCATGCCCATTTGGGGCTTCTCCCAAGAAATACAGTATGAGGAGC 255

QY 1794 ATTGAGGAAGAGATCTTAAGAGCATGATGTTGGCGTACAGAGAAATATCCTCCACCT 1853
    || || || || || || || || || || || || || || || || || || || ||
DB 256 CGTGACCTGCTGTGATGGGCTGTGATGATGATGAGTCCGAGCCGAGCTGATTTCCCAACTC 315

QY 1854 CTTCCGACCC 1863
    || || || || || || || || || || || || || || || || || || || ||
DB 316 CTTCCATGCC 325

RESULT 2
US-08-415-818-5
; Sequence 5, Application US/08415818
; Patent No. 5621079
; GENERAL INFORMATION:
```

```

1  APPLICANT: Cascieri, Margaret A.
2  APPLICANT: Linemeyer, David L.
3  APPLICANT: Macnell, Douglas J.
4  APPLICANT: Shiao, Lin-Lin
5  APPLICANT: Strader, Catherine D.
6  APPLICANT: Tan, Carina P.
7  APPLICANT: Weinberg, David H.
8  TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR
9  NUMBER OF SEQUENCES: 12
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Mary A. Appollina
12 STREET: P.O. Box 2000, 126 E. Lincoln Ave.
13 CITY: Rahway
14 STATE: NJ
15 COUNTRY: USA
16 ZIP: 07065
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.30
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/415,818
25 FILING DATE:
26 CLASSIFICATION: 435
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: 08/383,746
29 FILING DATE: 03-FEB-1995
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Appollina, Mary A.
32 REGISTRATION NUMBER: 34,087
33 REFERENCE/DOCKET NUMBER: 19390
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: 908-594-3462
36 TELEFAX: 908-594-4720
37 INFORMATION FOR SEQ ID NO: 5:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 2280 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: double
42 TOPOLOGY: linear
43 MOLECULE TYPE: DNA (genomic)
44 HYPOTHETICAL: NO
45 ANTI-SENSE: NO
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 822..1937
49 US-08-415-818-5

```

Query Match	Similarity	1.7%	Score 36.2	DB 1	Length 2280
Best Local	55.0%		Pred No. 0.35		
Matches	71	Conservative	0	Mismatches	56
				Indels	0
				Gaps	0
QY	409	GACAAAGTAGAAGACAGACTCAAGGGTATCACTGACGAGATGGAGCGTTCGCCAAGCT	468		
Db	501	GACAAAGGAGATCACAAGCTTAAAGCAATCATTTTGTACATAGTAGCTTTGAGGAAAGCT	560		
QY	469	CAGGAGGTATGCCAAGAACTGGCCAAAAGAAATGCAAGACGAGAACCACTCTGCATC	528		
Db	561	GAGGTACATCATGAACTCTCTCTCTCAAAAACAAAACAAAACAAAACCTCTAC	620		
QY	529	GTAGAGCCT	537		
Db	621	TAAATATCT	629		

RESULT3
US-08-894-236-5
; Sequence 5, Application US/08894236
; Patent No. 593263
; GENERAL INFORMATION:
; APPLICANT: Cascerfi, Margaret A
; APPLICANT: Linemeyer, David L.

```

1 APPLICANT: MacNeil, Douglas J.
2 APPLICANT: Shiao, Lin-Lin
3 APPLICANT: Strader, Catherine D.
4 APPLICANT: Tan, Carina P.
5 APPLICANT: Weinberg, David H.
6 TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR
7 NUMBER OF SEQUENCES: 14
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Mary A. Apollina
10 STREET: P.O. Box 2000, 126 E. Lincoln Ave.
11 CITY: Rahway
12 STATE: NJ
13 COUNTRY: USA
14 ZIP: 07065
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.30
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/894,236
22 FILING DATE:
23 CLASSIFICATION: 435
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: 08/283,746
26 FILING DATE: 03-FEB-1995
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: 08/415,818
29 FILING DATE: 03-APR-1995
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Apollina, Mary A.
32 REGISTRATION NUMBER: 34,087
33 REFERENCE/DOCKET NUMBER: 193907
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: 908-594-3462
36 TELEFAX: 908-594-4720
37 INFORMATION FOR SEQ ID NO: 5:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 2280 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: double
42 TOPOLOGY: linear
43 MOLECULE TYPE: DNA (genomic)
44 HYPOTHETICAL: NO
45 ANTI-SENSE: NO
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 822..1937
49 US-08-894-236-5

```

Query Match	1.7%	Score 36.2	DB 2	Length 2280
Best Local Similarity	55.0%	Pred. No. 0.35		
Matches 71	Conservative	0	Mismatches 58	Indels 0
			Gaps 0	
Qy	409 GACAAGTAGAAGAGACAGCTCAAGGGATACCTGACGAGATGAGACGCTCGCGCAAAGCT	468		
Db	501 GACAAGTAGAAGAGACAGCTTAAGGATCATCTTTGTACATAGTGAGTTGAGGAAAGCT	560		
Qy	469 CAGGAGGTATGCCAAGAACTGGCCAAAATAATGCAGACGAGACCACTCTGCATC	528		
Db	561 GAGGTACATCATGAACTCTCTCTCAAAAACAAAACAAAACAAACCTTTAC	620		
Qy	529 GTTAGAGCCT	537		
Db	621 TAATATCT	629		

```

RESULT 4
PCT-US96-01444-5
; Sequence 5, Application PC/TUS9601444
; GENERAL INFORMATION:
; APPLICANT: Cascieri, Margaret A.
; APPLICANT: Linemeyer, David L.

```



```

1 APPLICANT: Macnell, Douglas J.
2 APPLICANT: Shiao, Lin-Lin
3 APPLICANT: Strader, Catherine D.
4 APPLICANT: Tan, Carina P.
5 APPLICANT: Weinberg, David H.
6 TITLE OF INVENTION: NEUROPEPTIDE Y RECEPTOR
7 NUMBER OF SEQUENCES: 14
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: Mary A. Appollina
10 STREET: P.O. Box 2000, 126 E. Lincoln Ave.
11 CITY: Rahway
12 STATE: NJ
13 COUNTRY: USA
14 ZIP: 07065
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: Patentin Release #1.0, Version #1.30
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: PCT/US96/01444
22 FILING DATE:
23 CLASSIFICATION:
24 PRIOR APPLICATION DATA:
25 APPLICATION NUMBER: 08/383,746
26 FILING DATE: 03-FEB-1995
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: 08/415,818
29 FILING DATE: 03-APR-1995
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Appollina, Mary A.
32 REGISTRATION NUMBER: 34,087
33 REFERENCE/DOCKET NUMBER: 19390Y
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: 908-594-3462
36 TELEFAX: 908-594-4720
37 INFORMATION FOR SEQ ID NO: 5:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 2280 base pairs
40 TYPE: nucleic acid
41 STRANDEDNESS: double
42 TOPOLOGY: linear
43 MOLECULE TYPE: DNA (genomic)
44 HYPOTHEetical: NO
45 ANTI-SENSE: NO
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 822..1937
49 PCT-US96-01444-5
50
51 Query Match 1.7%; Score 36.2; DB 5; Length 2280;
52 Best Local Similarity 55.0%; Pred. No. 0.35;
53 Matches 71; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
54
55 Y 409 GACAAGGTAAAGACAGCTCAAGGCTATCACTGACGAGATGTGAGCGCTCTGGCCAAAGCT 468
56 DB 501 GACCAAGGAGGATCAGACAGCTTAAGACATCATTTGTACATAGTAGTGAGGAAAGCT 560
57 Y 469 CAGAGGATGATGCCAAGAACTGGCCAAAAGAAATGCAGACGAGAACCAACTGTGCATC 528
58 DB 561 GAGGTACATGACACTCTCTCTCTCAAAAACAAACAAACAAACAAACAAACCTTCTAC 620
59 Y 529 GTAGAGCCT 537
60 DB 621 TAATATCT 629
61
62 RESULT 5
63 US-08-232-463-14/c
64 Sequence 14, Application US/08232463
65 Patent No. 3670367
66 GENERAL INFORMATION:
67 APPLICANT: DORNER, F.

```

Db 1063 GCAAGCTCCTGACCTGCA 1044

RESULT 6

US-08-825-782-2

; Sequence 2, Application US/08825782

; Patent No. 5834239

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Goli, Surya K

; TITLE OF INVENTION: COFACTOR A-LIKE PROTEIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/825,782

; FILING DATE: Filed Herewith

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0263 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 449 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: PROSTUR08

; CLONE: 1654275

US-08-825-782-2

Query Match 1.6%; Score 35.2; DB 2; Length 449;

Best Local Similarity 53.7%; Pred. No. 0.25;

Matches 73; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 422 GACAGCTCAGGGTATCTACGACGATGAGCGTTCGCGAAGCTCAGAGGTATGCC 481

Db 89 GACAGATCAAGATCAAGACCGCGTGTGAGAGCGGTGTCAGAAAGAAAGTGAATG 148

QY 482 AAGAACTGGCCAAAAGAAATGACAGACGAGAACCACTGTCATGCTAGAGCTTTCT 541

Db 149 AAAAAGAGGCAAAACAAAGAAAGAAAGATTGAAGAAATGAGAGCTGAAGACGCTGAAA 208

QY 542 ATTCCACACAGCAATA 557

Db 209 ATTACGACATTAAAAA 224

RESULT 7

US-07-686-340-1/c

; Sequence 1, Application US/07686340

; Patent No. 5322785

; GENERAL INFORMATION:

; APPLICANT: Comb, Donald G.

; APPLICANT: Perler, Francine

; APPLICANT: Kucera, Rebecca

; APPLICANT: Jack, William E.

; TITLE OF INVENTION: Purified Thermostable DNA Polymerase

; TITLE OF INVENTION: Obtainable From Thermococcus litoralis

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS

; ADDRESSEE: 6 CUSHMAN

; STREET: 130 WATER STREET

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: US

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/686,340

; FILING DATE: 19910417

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/513,994

; FILING DATE: 26-APR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/626,057

; FILING DATE: 11-DEC-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: RESNICK, DAVID S.

; REGISTRATION NUMBER: 34235

; REFERENCE/DOCKET NUMBER: 39296-CIP-11

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 523-3400

; TELEFAX: (617) 523-6440

; TELEX: 200291 STRE UR

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 5837 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: double

; TOPOLOGY: linear

US-07-686-340-1

Query Match 1.6%; Score 34; DB 1; Length 5837;

Best Local Similarity 57.5%; Pred. No. 3.2;

Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1405 GTAATGACACAGAAATTCAGAGCGACACGAGAGAGATGACAGCTTGATTCGACAG 1464

Db 3686 GTCAATCGTCAGAGCTTAGTGCTTCAACACCTTCGAGATGAGATTCCTTTTCGCCAAT 3627

QY 1465 AGTGTGTTCTTATAGCGACAACTTCATGCCATGCTCAAGTCCA 1510

Db 3626 GCTGTAGTCCACCTTAGAAGAAAGATCTTATTTTCAAAATCTA 3581

RESULT 8

US-08-004-139B-1/c

; Sequence 1, Application US/08004139B

; Patent No. 5496714

; GENERAL INFORMATION:

; APPLICANT: Comb, DONALD G.

; APPLICANT: PERLER, FRANCINE B.

; APPLICANT: XU, MING-OUN

; APPLICANT: HODGES, ROBERT A.

; APPLICANT: JACK, WILLIAM E.

; TITLE OF INVENTION: MODIFIED PROTEINS AND METHODS OF THEIR

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &

; ADDRESSEE: CUSHMAN

; STREET: 130 WATER STREET

CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/004,139B
FILING DATE: 09-DEC-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 42188
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
US-08-004-139B-1

Query Match 1.6%; Score 34; DB 1; Length 5837;
Best Local Similarity 57.5%; Pred. No. 3.2;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 1405 GTCATGCAAGAATTCAAAGCGCAGACAGAGAGATGACACTTCGATTTCGACAG 1464
DB 3686 GTCATGTCGACAGATGAGCTTCACACACCTTCGACAGATGACATTCCTTTCCCAAT 3627
OY 1465 AGTGTGTTCCCTTATAGCGACAAGTTCATGCCATCGTCAGTCA 1510
DB 3626 GCTGTAGTCACCTTAGAGAAAAGATCCTTATTTCACAAATCTA 3581

RESULT 9
US-08-117-491-1/c
Sequence 1, Application US/08117491
GENERAL INFORMATION:
PATENT NO. 5500363
APPLICANT: Comb, Donald G.
APPLICANT: Perler, Francine
APPLICANT: Kucera, Rebecca
APPLICANT: Jack, William F.
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERS FROM ARCHAEABACTERIA
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,491
FILING DATE: 07-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/811,421

FILING DATE: 12-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/686,340
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057
FILING DATE: 11-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/513,994
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 39296C3FWC2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-117-491-1

Query Match 1.6%; Score 34; DB 1; Length 5837;
Best Local Similarity 57.5%; Pred. No. 3.2;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 1405 GTCATGCAAGAATTCAAAGCGCAGACAGAGAGATGACACTTCGATTTCGACAG 1464
DB 3686 GTCATGTCGACAGATGAGCTTCACACACCTTCGACAGATGACATTCCTTTCCCAAT 3627
OY 1465 AGTGTGTTCCCTTATAGCGACAAGTTCATGCCATCGTCAGTCA 1510
DB 3626 GCTGTAGTCACCTTAGAGAAAAGATCCTTATTTCACAAATCTA 3581

RESULT 10
US-08-271-364A-1/c
Sequence 1, Application US/08271364A
GENERAL INFORMATION:
PATENT NO. 5756334
APPLICANT: PERLER, FRANCINE B.
APPLICANT: SOUTHWORTH, MAURICE W.
TITLE OF INVENTION: RECOMBINANT THERMOSTABLE DNA POLYMERASE
TITLE OF INVENTION: FROM ARCHAEABACTERIA
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEW ENGLAND BIOLABS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/271,364A
FILING DATE: 06-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/811,421
FILING DATE: 18-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/686,340
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057
FILING DATE: 11-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/513,994
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Gregory D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-054C3FC2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-222-715B-1

Query Match
Best Local Similarity 1.6%; Score 34; DB 2; Length 5837;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1405 GTAATCGACAAGAAATTCAGAGCGACGACGAGAGTATGCACTTCTGTAATTCAGCAAG 1464
DB 3686 GGCATCGTCACAGTAGTGTCTTCAACACCTTCGAGAAATGCAATCTTTTTCGCCAAT 3627

QY 1465 AGTGTGCTCTTATAGCGACAGTTCATGCCATGCTCAAGTCCA 1510
DB 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAATCTA 3581

RESULT 13
PCT-US96-10545A-1/c
Sequence 1, Application PC/TUS9610545A
GENERAL INFORMATION:
APPLICANT: COMB, DONALD G.
APPLICANT: PERLER, FRANCINE B.
APPLICANT: JACK, WILLIAM E.
APPLICANT: XU, MING-QUN
APPLICANT: HODGES, ROBERT A.
APPLICANT: NOREN, CHRISTOPHER J.
TITLE OF INVENTION: MODIFIED PROTEINS AND METHODS OF THEIR
TITLE OF INVENTION: PRODUCTION
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:
ADDRESSEE: GREGORY D. WILLIAMS, NEW ENGLAND BIOLABS, INC.
STREET: 32 TOZER ROAD
CITY: BEVERLY
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 01915
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10545A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/580,555
FILING DATE: 29-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/496,247
FILING DATE: 28-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/146,885
FILING DATE: 03-NOV-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/004,139
FILING DATE: 09-DEC-1992
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-036C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5837 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
PCT-US96-10545A-1

Query Match
Best Local Similarity 1.6%; Score 34; DB 5; Length 5837;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1405 GTAATCGACAAGAAATTCAGAGCGACGACGAGAGTATGCACTTCTGTAATTCAGCAAG 1464
DB 3686 GGCATCGTCACAGTAGTGTCTTCAACACCTTCGAGAAATGCAATCTTTTTCGCCAAT 3627

QY 1465 AGTGTGCTCTTATAGCGACAGTTCATGCCATGCTCAAGTCCA 1510
DB 3626 GCTGTAGTCCACCTTAGAGAAAAGATCCTTTATTTTCACAATCTA 3581

RESULT 14
US-09-773-816-1
Sequence 1, Application US/09773816
Patent No. 6340774
GENERAL INFORMATION:
APPLICANT: Stanford University
APPLICANT: Khosla, Chaitan
TITLE OF INVENTION: NON-SPEROIDAL ESTROGEN-RECEPTOR
TITLE OF INVENTION: ANTAGONISTS
FILE REFERENCE: 28600-20210.00
CURRENT APPLICATION NUMBER: US/09/773,816
CURRENT FILING DATE: 2001-06-08
PRIOR APPLICATION NUMBER: US 60/243,458
PRIOR FILING DATE: 2000-10-25
PRIOR APPLICATION NUMBER: US 60/179,305
PRIOR FILING DATE: 2000-01-31
NUMBER OF SEQ ID NOS: 1
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 23673
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)-(23623)
OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

Query Match
Best Local Similarity 1.6%; Score 33.6; DB 4; Length 23673;
Matches 78; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 625 AAGTTCGAGCGGATACGACACTGATGTGCCCGCTCAGACGGCGACTTCAGCGTA 684
DB 12747 AAGTTCCTTCAGACGACCGCTGTGGGTGTTCCTCCGTCGCGACGCGCGTC 12806

QY 685 TTCCGCGTATAGCGGTCGACAAACGCGCGCGGAATACAGCAAGACATAAACCC 744

Db 12807 TTTCCGGGCTGTGCTACGCGTTGACCGCCCGCCGTTGCGCGGCGGACGCGCCCTCC 12866

745 TATAAGCCCGTTACTTCGCTGCCGATCCAT 776

Db 12867 TTCGAGGAGGTCACGACGCGGCCCGCACCAT 12898

RESULT 15

```

US-09-428-517-1
? Sequence 1, Application US/09428517
? Patent No. 6251636
? GENERAL INFORMATION:
? APPLICANT: Betlach, Mary C.
? APPLICANT: Shah, Sanjay Krishnakant
? APPLICANT: McDaniel, Robert
? APPLICANT: Tang, Li
? TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
? FILE REFERENCE: 30062-20029..00
? CURRENT APPLICATION NUMBER: US/09/428, 517
? CURRENT FILING DATE: 1999-10-28
? EARLIER APPLICATION NUMBER: 60/120, 254
? EARLIER FILING DATE: 1999-02-16
? EARLIER APPLICATION NUMBER: 60/106, 100
? EARLIER FILING DATE: 1998-10-29
? NUMBER OF SEQ ID NOS: 12
? SOFTWARE: Patentln Ver. 2.1
? SEQ ID NO 1
? LENGTH: 50937
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Recombinant DNA
US-09-428-517-1

```

Query Match	1.6%	Score	33.6	DB	4	Length	50937
Best Local Similarity	47.2%	Pred. NC	18				
Matches 102; Conservative	0	Mismatches	114	Indels	0	Gaps	0

QY	1377	CGACAAAGCTCCCCGATATATATTCACAAATATGTATATGCACAAAGAAATATTCAAAGCGACAGAA	1436
Db	44477	CGCCGAGATCTACGACGCCCGTGTACCCGGGGCCGGGCAAGACATACCCCGGGCAAGCGGAA	44536
QY	1437	GAAATATGACAGCTTCGTATTCGACAAAGAGTGATGGTTCTTATATAGCGACAAAGTTTCATCG	1496
Db	44537	GGACGTGGCGGACCTGTGCGGAGCCGGGATGGCCGAGCGGTCCTCTCTTGACAGTGGC	44596
QY	1497	CATGCTCAAGTCCATGGACAAAGAAAGTTTGCCAAAGCTATTCGAGAAAGATCCGGCACT	1556
Db	44597	CTGGGGACACGGGGCCACCTTCGGGACACTTGGCCACAGCTCTTCGACAGCAGCCCGCGGTCT	44656
QY	1557	AGAGCTTCCCAAGAGCGTAAATAGCTGCTGCTCGGCG	1592
Db	44657	CGAACTGTCCGCGAGCATGCTGTGACATGCGCCGCTC	44692

Search completed: October 17, 2002, 21:41:49
Job time : 243 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 17, 2002, 19:23:15 : Search time 1728 seconds
(without alignments)

16707.168 Million cell updates/sec

Title: US-10-008-355-1
Perfect score: 2139
Sequence: 1 atgcaatgaatnaaaag.....aagagctgagttgctctaa 2139

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estc2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pin:*
16: em_gss_vtl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	110	5.1	537	12	BH388186 AG-ND-124
C 2	110	5.1	591	12	BH392840 AG-ND-162
C 3	108.2	5.1	500	12	BH386991 AG-ND-148
C 4	103.8	4.9	418	12	BH375201 AG-ND-173
C 5	72.4	3.4	426	12	BH400391 AG-ND-147
C 6	66.4	3.1	697	12	BH375688 AG-ND-120
C 7	55	2.6	398	9	AW774807 EST333958
C 8	49.2	2.3	1005	12	BH164253 ENT76537F
C 9	42.8	2.0	555	12	BH400866 AG-ND-125
C 10	42.2	2.0	692	9	A1405287 GH25242.5
C 11	41.2	1.9	721	12	BH391233 AG-ND-141
C 12	39.4	1.8	822	12	BH371846 AG-ND-162
C 13	38.6	1.8	687	12	BH448937 BOC2L67R
C 14	38.4	1.8	184	9	AW375157 OV2-CT012
C 15	38.4	1.8	466	10	BF426788 d172a12.y
C 16	38.4	1.8	491	9	AW148212 d13h05.x
C 17	38.4	1.8	895	10	BF784434 602110820

18	38	1.8	465	9	AV628406
19	38	1.8	497	9	AV621707
20	38	1.8	587	10	B1247683
C 21	38	1.8	1091	10	BE282103
C 22	37.8	1.8	422	9	BB851067
C 23	37.8	1.8	461	9	AL372451
C 24	37.8	1.8	680	10	B1309279
C 25	37.8	1.8	773	10	B644863
C 26	37.6	1.8	422	9	AV629946
27	37.6	1.8	927	10	BG437679
28	37.4	1.7	367	10	BG983576
29	37.2	1.7	318	9	AV644201
30	37.2	1.7	348	10	BE717547
31	36.8	1.7	668	10	AU070770
32	36.6	1.7	313	10	BE804382
33	36.6	1.7	436	9	AV637236
34	36.6	1.7	471	9	AV640279
35	36.6	1.7	478	9	AV643425
36	36.6	1.7	497	10	BE552689
37	36.6	1.7	509	10	BE252639
38	36.6	1.7	517	9	AV388690
39	36.6	1.7	517	9	AV635803
40	36.6	1.7	552	10	B1643498
41	36.6	1.7	586	10	B1675086
42	36.4	1.7	409	9	AA277303
43	36.4	1.7	459	9	AW154047
44	36.4	1.7	482	9	AA541836
C 45	36.4	1.7	515	9	AV743237

ALIGNMENTS

RESULT 1
LOCUS BH388186/c
DEFINITION AG-ND-124P3.TF ND-TAM Anopheles gambiae genomic clone AG-ND-124P3,
DNA sequence.
ACCESSION BH388186
VERSION BH388186.1 GI:17334327
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anophelinae.
REFERENCE 1 (bases 1 to 537)
AUTHORS Shetty J., Malek J., Koo H., Collins F., Gardner M. and Loftus B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-124P3.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1..537
/organism="Anopheles gambiae"
/strain="PEST"

FEATURES
source

```

/bd_xref="taxon:7165"
/clone="AG-ND-124P3"
/clone_11b="ND-TAM"
/notes="Vector: pECBAC1; Site_1: HindIII"

```

Query Match	5.1%	Score 110	DB 12	Length 537
Best Local Similarity	55.9%	Pred. No. 1.2e-19		
Matches 209	Conservative 0	Mismatches 165	Indels 0	Gaps 0

[illegible]

RESULT 2	
BH392840/c	
LOCUS	591 bp DNA linear GSS 11-DEC-2001
DEFINITION	AG-ND-162P14.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162P14
ACCESSION	BH392840
VERSION	BH392840
KEYWORDS	' DNA sequence.
SOURCE	BH392840.1 GI:17338981
ORGANISM	GSS. African malaria mosquito.
REFERENCE	Anopheles gambiae Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae ; Anophelinae. 1 (bases 1 to 591) Shetty,J., Malek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J Direct Submission of BAC-end sequences from Anopheles gambiae Unpublished (2001)
COMMENT	Other_GSSs: AG-ND-162P14.TR

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org

This clone is from an *A. gambiae* BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII

```

partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1. 591
source
FEATURES

```

```

/NOTE="Vector: PECBAC1; Site_1: HANDIII"
BASE COUNT      170 a      122 c      105 g      194 t
ORIGIN

```

	Query Match	5.1%;	Score 110;	DB 12;	Length 591.
	Best Local Similarity	55.9%;	Pred. No. 1,2e-15;		
	Matches 209;	Conservative 0;	Mismatches 165;	Indels 0;	Gaps 0;
OY	527 TCGTAGACCCCTTTCATTATCCAAACAAGAAACTCTCCATCGCTCAGATGATTAATCAAG	586			
Db	374 TGGTGAAGATTTCTTCAAAGAAATAGATTTACTACTTTGGTATTCAGATTTCAAG	315			
OY	587 ACGTTCCGATGATTAATTCCTCCTCCACCCTCGAATGATACTGGAGGCATACCGGCA	646			
Db	314 ATGTTCGTGTAGATACCTCCGCCGTCACTATATGGTAATATGATGATATGTATG	255			
OY	647 ACTGATATGCGCCGCGTCACACGAGCGACTTCAGCGTATTCGCCGTATGCCGTGCCG	706			
Db	254 ACTGGGAATGGCGCAAGACACACAGAGAGACTTCTGTGTTTTCCGTGTTATGCTGATMAA	195			
OY	707 ACAACGGCCGGCGCATATACAGCAAGGACAAATAAACCTATATAGCCGTTTACTTGCGTG	766			
Db	194 ATGGCAATCTCGTGTAATATTTCAAGTAAACAGCTTCTTAAAGCCTTAAGCATCATTTTAC	135			
OY	767 CCGTATCCATFCAAGAGGCTACAAAGCGTGAAGCATATGACATACCATCGGTTTCCGGGCA	826			
Db	134 CAATTTCACTAAAGAATATAAGCCTGGTGATTTTGGCAATATATTTAGATATCTCGSTA	75			
OY	827 GTACGATCGCTACCTCACTTCTTGGGGTGTGGAAAGATCGATTCGAAAACGAACAATC	886			
Db	74 CAACAAACAGATACCTTACTTCTTCGGAATTCGAACAATATGTAAGCAAAAGATTACCCGG	15			
OY	887 CTCGTATGAGATT 900				
Db	14 CATGGCTTGAAGCT 1				

RESULT 3					
LOCUS	BH386991/c				
DEFINITION	BH386991	500 bp	DNA	linear	GSS 10-DEC-2001
ACCSSION	AG-ND-148L11	TR	ND-TAM	Anopheles gambiae	genomic clone AG-ND-148L11
VERSION	' DNA sequence.				
KEYWORDS	BH386991				
SOURCE	BH386991.1	GI:17331133			
ORGANISM	GSS.				
	African malaria mosquito.				
	Anopheles gambiae				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
	Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae				
	; Anophelae.				
REFERENCE	1 (bases 1 to 500)				
AUTHORS	Shetty,J., Malek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.				
TITLE	Direct Submission of BAC end sequences from Anopheles gambiae				
JOURNAL	Unpublished (2001)				
COMMENT	Other_GSSs: AG-ND-148L11.TF				

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 426)
Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M., and Loftus, B. J
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)

Euarystota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoides
Anophelinae.

COMMENT
Other_GSSs: AG-ND-147H4.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@ligr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
source 1..426
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-147H4"
/clone_11b="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 111 a 79 c 76 g 160 t
ORIGIN

Query Match 3.4%; Score 72.4; DB 12; Length 426;
Best Local Similarity 57.8%; Pred. No. 3.6e-09;
Matches 149; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 1756 GGTGCTGTACACATATATACAGAGGCGCTATTGGAGACGATCCTAAG 1815
|||||
DB 258 GGTATACACAGATTATACCAATGAGGATATTAAGATACAGAAAGGT 199
QY 1816 AGCATGATTTGGCCGTACAGAGAAATATCTCGACCTTCGCGCAAAAATATGGT 1875
|||||
DB 198 GACGAAGATTCGATCTTCACAGACCTTGTGATCTTTATTAAGAAAAATTTGGT 139
QY 1876 ---CGCTATGCGAGAGCGTCACATTCATGCTTTCTTATCGAACAAGCATCAG 1932
|||||
DB 138 ATGTACAAAGACAAAGACGGGCACTTCATGTAACCTCTCTTCTAATACGATTTACA 79
QY 1933 GGGGTAACTCCGATAGCCCGGATTCGATTAAGAGCGCCGCTGATCGTCTTTC 1992
|||||
DB 78 GGAGTAACCTCAAGTCTCAATTAATGATGTTACGGAAGACTTATAGTCTTGCATTT 19
QY 1993 GATGCACTGGGAGCT 2010
|||
DB 18 GACGGAAACAGTGAAGCT 1

RESULT 6
BH375688/c 697 bp DNA linear GSS 10-DEC-2001
LOCUS AG-ND-120J9.TR ND-TAM Anopheles gambiae genomic clone AG-ND-120J9,
DEFINITION DNA sequence.
ACCESSION BH375688
VERSION BH375688
KEYWORDS BH375688.1 GI:17321830
SOURCE GSS.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
REFERENCE 1 (bases 1 to 697)
AUTHORS Shetty J., Malek J., Koo H., Collins F., Gardner M. and Loftus B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-120J9.TF

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@ligr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers
source 1..697
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-120J9"
/clone_11b="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 214 a 133 c 122 g 228 t
ORIGIN

Query Match 3.1%; Score 66.4; DB 12; Length 697;
Best Local Similarity 50.3%; Pred. No. 1.2e-07;
Matches 197; Conservative 0; Mismatches 186; Indels 9; Gaps 1;

QY 16 AAAGATATCTCTGGAGCAGCCCGCTGTGGGTCTTCAGGGGTAGCCAAAGCCGAC 75
|||||
DB 475 AAAAATCTCTATTAGCACTACCTTACTCCCGAGAGCGCTTTGCTCAACAGCGG 416
QY 76 AAAGCATGTGGCTCCTCAACGACTCATCAGAGAAATCTGATCGATCGGTAGCTC 135
|||||
DB 415 GGAGATGTGATTCCTACAGATTAATGAAGAA-----AATGAAGAAATTG 365
QY 136 GCGTTAGCGTCCCGTGGATTCCTACAGTTTCGACACAGCCGTCATTCGCAATGCC 195
|||||
DB 364 GCATGAAAGATTTTCGCGAAGATATTTCAATACTCAAAAACCTAATTAAGATGCT 305
QY 196 GTGGTTATCTGGGTGGCGGATGACCGGTATCAGAGTCCGATCAGGCGCTGATCTT 255
|||||
DB 304 GTAGTCAATTCACAGGAGCGTGTACTGTAATTAATTTCTCAAAAAGTTGTTATTG 245
QY 256 ACCAACAACCACTGGCGATACGTCCTATCCAGAGCCAAAGCAGGTGATCAGACTAT 315
|||||
DB 244 ACAAAATCACCACTGTGGCTTCGACAGATTCAGCTCANTGCATGTCANAAATATACCT 185
QY 316 CTGCGGATGTTTGGTTCGCAAGATGGGTGAGAGACTTCGATTCGGGCTTTCC 375
|||||
DB 184 CTAAATGATGTTTCTGGGCAAAAAACATGGGAGAAACCTTCAAAACCTGTTGTT 125
QY 376 GTGAATATCTGCGCAAGATCTGAAGTAAAC 407
|||
DB 124 GTAGATTTCATCACCAGATTAATTAAGAGGTGAC 93

RESULT 7
AW774807 398 bp mRNA linear EST 07-SEP-2000
LOCUS AW774807
DEFINITION EST333958 KV3 Medicago truncatula cDNA clone PKV3-24C3, mRNA
sequence.
ACCESSION AW774807
VERSION AW774807
KEYWORDS AW774807.1 GI:7718724
SOURCE EST.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;

REFERENCE
1 (bases 1 to 398)

Vandenbosch,K., Hurt,J., Moore,J., Beremand,P., Peng,H., Ellis,L.,
Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and
Fraser,C.M.

ESTs from roots of Medicago truncatula after Rhizobium inoculation
Unpublished (1999)

JOURNAL

COMMENT

Contact: Vandenbosch K
Department of Biology
Texas A&M University
College Station, TX 77843-3258, USA
Tel: 409 845 2797
Fax: 409 845 2891

Email: Kate@mail.bio.tamu.edu
Texas A&M EST name: T258226e

TIGR sequence name: MTEBG14TK

More information is available at:

http://chrysis.tamu.edu/medicago

Seq primer: Skmod (CTA gaa CTA gtc gat CC).

FEATURES

source

Location/Qualifiers

1..398

/organism="Medicago truncatula"

/cultivar="genotype A17"

/db_xref="taxon:3880"

/clone="PKV3-24C3"

/clone_1db="KV3"

/tissue_type="Seedling roots"

/dev_stage="3 days post-inoculation with Sinorhizobium

meliloti"

/lab_host="E. coli strain XLOLR"

/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:

XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA

was directionally ligated into the unizap XR vector from

stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-zap phage using Ex-assist

helper phage and propagated in XLOLR cells."

BASE COUNT 137 a 64 c 81 g 116 t

ORIGIN

Query Match 2.6%; Score 55; DB 9; Length 398;

Best Local Similarity 50.6%; Pred. No. 0.00026;

Matches 133; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 67 AAGCCGACAAAGCGATGTCCTCAAGCACTCAATCAGAGAACTGGATCGAATG 126

Db 28 AGACCCGATGAAAGAAATGCTGCTCTGATGTTATCGAAAGATTAAATCACAGGATATG 87

QY 127 CGTAGCTCGGCTTTAGGCTCCCTTGATTCGCTACAGTTTCGACAAAGCCGCTCAT 186

Db 88 CAAAATAATGGCTTGCAATTCACAGCCGAAAGAAATTTACACATTACCAATCACAGCTTA 147

QY 187 GCCAATGCCGCTTATCTTCTGCGGCGATGACCGGATACAGATGTCGATCAGAGCC 246

Db 148 AAAATGCGATGTACATTTCAATGAGGCTGTACTGCAAGAAATCGTTTCAAAAGCGGT 207

QY 247 CTGATCTTTACCAACCACTGCGGATACGGTCTATCCAGACCAAGACGGTGTGAT 306

Db 208 TTGGTCTTACCAATCACCACTGTGTTATAGCCGATTCGAGAACTTCAACTGACGAA 267

QY 307 CACGACTATCGCGGATGTTT 329

Db 268 CAAAATCATTTTAAAAATGGTTT 290

RESULT 8

LOCUS BH164253/c

DEFINITION BH164253 1005 bp DNA linear GSS 24-SEP-2001

SOURCE ENT637P Entamoeba histolytica Sheared DNA Entamoeba histolytica

ACCESSION BH164253

VERSION BH164253.1 GI:15737691

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 1005)

Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.

Determination of clone end sequences from Entamoeba histolytica

HMI:IMSS sheared DNA library (2001)

JOURNAL

COMMENT

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

Email: b1loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: Shotgun

High quality sequence start: 15

High quality sequence stop: 487.

FEATURES

source

Location/Qualifiers

1..1005

/organism="Entamoeba histolytica"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_1db="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOSI; Site_1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

light size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barelli, Oxford University Press, 1999)."

BASE COUNT 186 a 314 c 336 g 169 t

ORIGIN

Query Match 2.3%; Score 49.2; DB 12; Length 1005;

Best Local Similarity 56.0%; Pred. No. 0.018;

Matches 93; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 195 CGGTGTTATCTTGGTGGCGGATGATCCGATACAGTGTCCGATCAGGCGCTGATCTT 254

Db 229 CGCGGTGGTGGCGCGCGGTGTGATCCGTTCTTCTGTCCGCGCGCGCTGCTGT 170

QY 255 TACCAACACACATGCGGATAGGTGATATCCAGAGCCAAAGCAGGATCAGACTA 314

Db 169 GACCAACACACATGCGGATAGGTGATATCCAGATACAAAGCTGCCCGACAGACACT 110

QY 315 TCTGCGCGATGTTGCTTCTCCAGAGATGGTGTAGAGGCTTCGG 360

Db 109 GATCAGTGTGGTCTTCATCCAGCGCGCTGCGAGAGCGCCCG 64

RESULT 9

LOCUS BH400866/c

DEFINITION BH400866 555 bp DNA linear GSS 11-DEC-2001

SOURCE AG-ND-125M4.TR ND-TAM Anopheles gambiae genomic clone AG-ND-125M4,

DNA sequence.

ACCESSION BH400866.1 GI:17347082

VERSION BH400866

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Class: BAC ends.
Location/Qualifiers

FEATURES
source 1..721
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-141L2"
/clone.lib="ND-TAM"
/note="Vector: pECBAC1; Site:1: HindIII"

BASE COUNT 269 a 122 c 147 g 183 t
ORIGIN

Query Match 1.9%; Score 41.2; DB 12; Length 721;
Best Local Similarity 47.3%; Pred. No. 2.6; Mismatches 138; Indels 0; Gaps 0;
Matches 124; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 924 GAAGAAGCCATGAGCGAGATCAGGCTACCCGTATCAATATGCGAGCAAGTATGCTCA 983
|||||
DB 28 GAAGAAGCATATGATTAAGACGATGCTACAAAGATTAGCTTACGCGCTAACTATGCAAG 87
|||||

QY 984 GAGTGTAAGTATGGAAGATTCGATTCGATGAAACCGCGCTCGCTCTTGACGT 1043
|||||

DB 88 TGTGTAAGTATGGAAGATTCGATTCGATGAAACCGCGCTCGCTCTTGACGT 147
|||||

QY 1044 GATAGGTCGTAAGCGTCCGAGAGAAAGACATTCGAGACTGATCCGTAAAGACGCA 1103
|||||

DB 148 AATGTTGACAAAAAAGAGATTGAGAAAAATTCACAACTAGCGCTGATTAAGCTGAAA 207
|||||

QY 1104 GAGTGTGCTATGCGGATGATTTCTCTCGAAAAGGCTTATPAGAGAGAGCCAA 1163
|||||

DB 208 CAAGCGCTTTATGGAATGATTAACCAATCTGATGCTACTACAGCAAAATTTCTAA 267
|||||

QY 1164 GGCCACCGTGAGTACTTAT 1185
|||||

DB 268 CAGAAATATCGAGAAAAATTAT 289
|||||

RESULT 12
BH371846 822 bp DNA linear GSS 10-DEC-2001
LOCUS AG-ND-162M17.TF ND-TAM Anopheles gambiae genomic clone AG-ND-162M17
DEFINITION , DNA sequence.
ACCESSION BH371846
VERSION BH371846.1 GI:17317971
KEYWORDS GSS
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
1 (bases 1 to 822)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other GSSs: AG-ND-162M17.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
Seq primer: M13 For
Class: BAC ends.

FEATURES
source 1..822
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-162M17"
/clone.lib="ND-TAM"
/note="Vector: pECBAC1; Site:1: HindIII"

BASE COUNT 235 a 142 c 155 g 290 t
ORIGIN

Query Match 1.8%; Score 39.4; DB 12; Length 822;
Best Local Similarity 65.2%; Pred. No. 9; Mismatches 31; Indels 0; Gaps 0;
Matches 58; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2007 AGCTATGATGTTGACATCGAGTTCGAAACCGATTCGACCGACATACGCTGACAT 2066
|||||

DB 1 ACCTTAAAGCGGTGACATTTGTTTGAACCTTAATTAAGAAAGACGATTAACTAGACGT 60
|||||

QY 2067 CCGCTACGTTCTCTCATGATTCGACAAAT 2095
|||||

DB 61 TAGATACGTAAGTTGGGTATGACAAAT 89
|||||

RESULT 13
BH448937 687 bp DNA linear GSS 12-DEC-2001
LOCUS BOGZL67TR BOGZ Brassica oleracea genomic clone BOGZL67, DNA
DEFINITION sequence.
ACCESSION BH448937
VERSION BH448937.1 GI:17634648
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 687)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other GSSs: BOGZL67TF
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers

FEATURES
source 1..687
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOGZL67"
/clone.lib="BOGZ"
/note="Vector: pHO51; Site:1: BstXI; 2-3 kb sheared
genomic DNA inserted into pHO51 using BstXI linkers"

BASE COUNT 209 a 117 c 159 g 202 t
ORIGIN

Query Match 1.8%; Score 38.6; DB 12; Length 687;
Best Local Similarity 53.7%; Pred. No. 14; Mismatches 69; Indels 0; Gaps 0;
Matches 80; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1457 TCACAGAAGTGTGGTTCCTTATAGGAGCAAGTTCATGCGATGCTCAATCCATGACA 1516
|||||

DB 685 TCACACCTTTTGTGGATCAAAAGACGACGATTGAGATCAAGATCTAGTTAAGC 626
|||||

QY 1517 AGGAAAGTTTGGCAAGGCTATCGAGAAAGATCCGACAGTAGAGCTTCCAAAGAGGTAA 1576
|||||

Db	625	AGTGAACAATGGCCATCGCTATACCTGATTTAGGAGTTGAGCTCTAATAATATCATGC	566
OY	1577	TAGCGTCGCTCGCCCTATTCAGGCCGAT	1605
Db	565	ATGCACCTGTTAGCGCTCTATAGCATTTAT	537
RESULT 14			
LOCUS	AM375157		
DEFINITION	AM375157	184 bp	mRNA
ACCESSION	AM375157	04-CT0121	linear
VERSION	AM375157.1	GI:6879811	EST
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 184)		
JOURNAL	HCCP http://www.ludwig.org.br/ORESTES .		
COMMENT	The FAPESP/LICR Human Cancer Genome Project		
	Unpublished (1999)		
	Contact: Simpson A.J.G.		

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: aslmpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?lvt1-QV2&t2-QV2-CT0121-1280999-001-e01&t3-1999-09-28&t4-1>)
Seq primer: puc 18 forward
High quality sequence start: 18
High quality sequence stop: 88.

```

FEATURES
source
    Location/Qualifiers
        1..184
            /organism="Homo sapiens"
            /db_xref="taxon:3606"
            /clone_lib="CR0121"
            /dev_stage="Adult"
            /note="Organ: colon; Vector: puc18; Site.1: Sma1; Site.2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent Application No. 1967,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT
29 a 53 c 58 g 44 t
ORIGIN

```

Query Match	1.8%	Score 38.4	DB 9	Length 184
Best Local Similarity	51.1%	Pred. No. 7.9		
Matches	90	Conservative	0	Mismatches 86
				Indels 0
				Gaps 0
OY	612	CAGCTCTGTAGTAGTTCGAGCGGATACGACAACTGATGTGGCCGCTCACACGGG	671	
Db	8	CAGCTTTGTCCTTCAGCTGCTTGGGCTTCTGGAACTTGGGGTGGGGCCCTGCTCGAT	67	
OY	672	GCACCTTCAGCGGATATCCGCGTATGCGGGTCCGACAAACCGCGCGCGGATACAGCA	731	
Db	68	CACCTTCTCCACGAAGCTCTTGTCCGTGGCTTTGGGGAAACGACACTCTCGTCCAGCA	127	
OY	732	GGACATATAACCCGATTAAGCCGTTTACTTGCCTCCGATATCATGCAAGGCTACA	787	
Db	128	GGCCGAATGCCCGGGGGGCGCTGCTGCTTCTCAATGAAGTGTGATGCAAGGGCTGCA	183	

RESULT 15	
LOCUS	EST 18-APR-2001
BF426788	466 bp mRNA 1 linear

DEFINITION	dff2a12.y1 Xenopus laevis unfertilized egg cDNA library Xenopus laevis cDNA clone IMAGE:3744694.5' similar to TR:O93308 O93308.1S
ACCESSION	BF426788
VERSION	BF426788.1 GI:11438250
KEYWORDS	EST.
SOURCE	African clawed frog.
ORGANISM	Xenopus laevis
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae; Xenopodinae; Xenopus. 1 (bases 1 to 466)
AUTHORS	Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Page,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
TITLE	Washu Xenopus EST project. 1999
JOURNAL	Unpublished (1999)
COMMENT	Contact: Sandy Clifton, Ph.D. University of Washington, Seattle, WA 98195-5060

Washu Xenopus esi project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: <http://www.resgen.com/> Please reference the id listed
below when ordering this clone: Source lab clone id - xlnmeg006b23
Seq primer: -40NP from G1bco
High quality sequence stop: 407.

FEATURES	SOURCE
location/Qualifiers	1..466
/organism="Xenopus laevis"	
/db_xref="taxon:8335"	
/clone="IMAGE:3744694"	
/clone_1lb="Xenopus laevis unfertilized egg cDNA library"	
/tissue-type="unfertilized egg"	
/lab_host="Top-10 F"	
/note="Vector: plasmid SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2ug of poly A+ RNA. EcoRI-XhoI cut cDNA was then ligated into Unizap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by	

PCR) to cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electrotransformed into Top-10 F'. Original library construction by Bruce Blumberg (Blumberg et al., 1991 Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9, 2923-2935). Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT	184 a	80 c	131 g	71 t
ORIGIN				
Query Match		1.8%;	Score 38.4;	DB 10; Length 466;
Best Local Similarity		56.2%;	Pred. No. 13;	
Matches	72;	Conservative	0;	Mismatches 56; Indels 0; Gaps 0;

[illegible]

Fri Oct 18,08:16:51 2002

us-10-008-355-1.rst

Page 9

Db 225 CTGAATCA 232

Search completed: October 17, 2002, 21:37:38
Job time : 1743 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 21:35:21 ; Search time 60 Seconds
(without alignments)
1318.076 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719
Sequence: 1 MWMKLSILIGALLGASG.....LFMDKMGCCPRLDGLKLI 712

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	133	3.6	883	22	AAU37789
2	130	3.5	883	20	AAU08339
3	130	3.5	883	21	AAU90514
4	127.5	3.4	1194	18	AAW21725
5	127	3.4	883	22	AAU38091
6	123.5	3.3	713	20	AAU43380
7	122.5	3.3	800	18	AAW21723
8	122.5	3.3	813	18	AAW21728
9	122.5	3.3	1181	18	AAW21727
10	122.5	3.3	1194	18	AAW21724
11	122.5	3.3	1194	18	AAW21726

12	117.5	3.2	951	20	AAV34536	Porphyrinoma gng
13	117.5	3.2	953	20	AAV34403	Porphyrinoma gng
14	114	3.1	711	17	AA88649	Neisseria meningit
15	113	3.0	708	11	AA843567	Human cancer assoc
16	113	3.0	2042	19	AAW56319	Haemophilus paraga
17	112.5	3.0	907	22	AA678650	XRN-100, unidentified
18	112	3.0	724	22	ABG15384	Novel human diago
19	112	3.0	724	22	ABG17531	Novel human diago
20	110.5	3.0	1027	14	AA842203	Protein L. Peptoc
21	110.5	3.0	1027	14	AA843699	Protein L. Peptoc
22	110	3.0	705	17	AA88645	Neisseria meningit
23	109.5	2.9	845	21	AA90942	Cenarchaeum symbio
24	109	2.9	416	21	AA834432	Gene 46 human secr
25	109	2.9	416	21	AA834433	Human secreted pro
26	109	2.9	1577	17	AA891047	Alpha-D-glucosyltr
27	109	2.9	2234	21	AAV81502	Streptococcus pneu
28	108.5	2.9	950	21	AA842742	Human ORF2506
29	108.5	2.9	950	22	AA893087	Human protein sequ
30	108	2.9	2027	22	ABG07898	Novel human diago
31	108	2.9	4536	20	AA896826	Amno acid sequenc
32	107.5	2.9	655	20	AAV49226	Cbpa of serotype 4
33	107.5	2.9	655	20	AAV32099	Choline binding pr
34	107.5	2.9	694	21	AA81653	Streptococcus pneu
35	106.5	2.9	523	22	AAU36325	Pseudomonas aerugi
36	106.5	2.9	1162	22	ABG09158	Novel human diago
37	106.5	2.9	1644	22	ABG12176	Novel human diago
38	106	2.9	918	22	AB859819	Drosophila melanog
39	106	2.9	912	14	AA851281	Helminth aminopept
40	106	2.9	1592	14	AA832925	Glucosyltransfera
41	106	2.9	2721	8	AA870647	Sequence of N-term
42	105.5	2.8	403	14	AA840855	38kd regression as
43	105.5	2.8	609	17	AA886557	N.meningitidis B21
44	105.5	2.8	710	22	AB862130	Drosophila melanog
45	105	2.8	361	22	AA860881	Mevalonate pathway

ALIGNMENTS

RESULT 1

AAU37789 standard; Protein: 883 AA.

AAU37789:

14-FEB-2002 (first entry)

Streptococcus pneumoniae cellular proliferation protein #218.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

OS Streptococcus pneumoniae.

PN W0200170955-A2.

XX 27-SEP-2001.

PD 21-MAR-2001; 2001WC-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,

PI Yamamoto RT, Xu HH;

PI

DR WPI: 2001-611495/70.
DR N-PSDB: AAS55648.

PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 13382; 511pp; English.

XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
XX Sequence 883 AA:

Query Match 3.6%; Score 133; DB 22; Length 883;
Best Local Similarity 20.4%; Pred. No. 0.025;
Matches 104; Conservative 65; Mismatches 166; Indels 176; Gaps 24;

QY 192 DVF-----KDVRYVFPSSVCGFGGDTNNMMPRTTGFVSFRVYAGADNRPAREYS 243
DB 26 DVFKPSGQOKAPYSIVPPNVTGKL-----HLG-----HA 57
QY 244 KDNKPYKPYFAVAVSMOGKADYAMTIGFPGSTDRYL-TSMGVEDRIENNNPREVNG 302
DB 58 WDT-TLQDIIIRQKRMGFD-----TLMLPGMDHAGINTQAKVEERLNGEGISRDLGR 110
QY 303 -----IKQIGWKEMASDAQTRIKYA-SKYAQSANTWKNKSGNNGIARLDVIGRRRAER 357
DB 111 ESFLTRKWMEMKD-----EYATYTIKEOMGKMGSLVDYSRERFTLDEGLS-----KAVRK 158
QY 358 APADMIRK-----NGKSAVYGD-----VLSTLEKAYK-----EGAKANREMT 394
DB 159 VVDLYKKKMWIRGEFTIIMWDPARTALSDIEVIHKDVGAFYHNNYMLEDSRALEVAT 218
QY 395 VLSETLFGTEVVRFAQFANALATNPDAHAGILKSLDKYKDYLS-----LDRKVLPM 449
DB 219 TRPETMFGDV-----AIAVNPDE-----DPRYKDLIGKNVLLPANKLIPIV 259
QY 450 LD-----IVRRIRPADKLPDIFKKNVIDKKFKGDTKKYADVFEDKSVVPYSDK 496
DB 260 GDEHADPERGTGVKITTPAHDPNDPLVGQRHMLPOVNNMNDGTMMNLEFESG-----DR 316
QY 497 FHAMLKSMDEKFAKAIEKDPAVELSKSYIAARA-----IQAD 535
DB 317 FEAR-----KAVYAKLEELGALVKIEKRVHSVGHSERTGVVVEPRLSITQWFKMQDLAKN 371
QY 536 AANANVAIEKGRLEFAGIREMYPGRALPS-----DANFTMMSYGSIKGYEPDQG 586
DB 372 AIANQDTEKK-----VEFYPRFNDTFLQWMMENVDWVISRQLMNGH-----QIP 416
QY 587 AANYNHTTGGKLVLEKODPKSDEPAVOENITLD 617
DB 417 AWMN--ADGEMTVGEAREPGDGTODEVDLD 445

RESULT 2
AAV08339

ID AAV08339 standard; Protein; 883 AA.

XX AAV08339;

AC 19-JUL-1999 (first entry)

DE S. pneumoniae vals protein.

KW vals; treatment; antagonist.

OS Streptococcus pneumoniae.

PN JP11113577-A.

PD 27-APR-1999.

PF 17-OCT-1997; 97JP-0321886.

PR 17-OCT-1997; 97JP-0321886.

PA (SMIK) SMITHKLINE BEECHAM CORP.

DR WPI: 1999-320830/27.

DR N-PSDB: AAX57027.

PT New vals polypeptide and polynucleotide - useful for treatment of
XX conditions associated with abnormal vals protein levels

PS Claim 1a; Page 25-27; 27pp; Japanese.

XX
XX This invention describes the isolation of a novel Streptococcus
CC pneumoniae vals protein. The new polypeptide and antagonist are
CC useful for the treatment of an individual with abnormal vals
CC protein levels, by administering vals to individuals requiring
CC vals polypeptide, and administering the antagonist to individuals
CC requiring inhibition of vals polypeptide.

XX
XX Sequence 883 AA:

Query Match 3.5%; Score 130; DB 20; Length 883;
Best Local Similarity 20.4%; Pred. No. 0.046;
Matches 104; Conservative 63; Mismatches 168; Indels 176; Gaps 24;

QY 192 DVF-----KDVRYVFPSSVCGFGGDTNNMMPRTTGFVSFRVYAGADNRPAREYS 243
DB 26 DVFKPSGQOKAPYSIVPPNVTGKL-----HLG-----HA 57
QY 244 KDNKPYKPYFAVAVSMOGKADYAMTIGFPGSTDRYL-TSMGVEDRIENNNPREVNG 302
DB 58 WDT-TLQDIIIRQKRMGFD-----TLMLPGMDHAGINTQAKVEERLNGEGISRDLGR 110
QY 303 -----IKQIGWKEMASDAQTRIKYA-SKYAQSANTWKNKSGNNGIARLDVIGRRRAER 357
DB 111 ESFLTRKWMEMKD-----EYATYTIKEOMGKMGSLVDYSRERFTLDEGLS-----KAVRK 158
QY 358 APADMIRK-----NGKSAVYGD-----VLSTLEKAYK-----EGAKANREMT 396
DB 159 VVDLYKKKMWIRGEFTIIMWDPARTALSDIEVIHKDVGAFYHNNYMLEDSRALEVAT 218
QY 397 --SETLFGTEVVRFAQFANALATNPDAHAGILKSLDKYKDYLS-----LDRKVLPM 449
DB 219 TRPETMFGDV-----AIAVNPDE-----DPRYKDLIGKNVLLPANKLIPIV 259
QY 450 LD-----IVRRIRPADKLPDIFKKNVIDKKFKGDTKKYADVFEDKSVVPYSDK 496
DB 260 GDEHADPERGTGVKITTPAHDPNDPLVGQRHMLPOVNNMNDGTMMNLEFESG-----DR 316
QY 497 FHAMLKSMDEKFAKAIEKDPAVELSKSYIAARA-----IQAD 535
DB 317 FEAR-----KAVYAKLEELGALVKIEKRVHSVGHSERTGVVVEPRLSITQWFKMQDLAKN 371
QY 536 AANANVAIEKGRLEFAGIREMYPGRALPS-----DANFTMMSYGSIKGYEPDQG 586

DB 372 AIANODEDK-----VEFYPPRFNDFTLQWMEVHDWVISRLQWNGH-----QIP 416

QY 587 AMYNHTTGKGVLEKODPKSDEFAVOENIID 617

DB 417 AMYN--ADGEMVYGEBAPEGDGTODEDYLD 445

RESULT 3

AA90514

ID AA90514 standard; Protein; 883 AA.

AC AA90514;

XX

DT 15-AUG-2000 (first entry)

DE Streptococcus pneumoniae valyl tRNA synthetase (vals).

XX

KM Valyl tRNA synthetase; vals; inhibitor; stringent response;

KW drug screening; antibacterial; antibiotic; genetic immunisation;

KM antibody; bacterial infection; meningitis.

XX

OS Streptococcus pneumoniae strain 0100993.

XX

PN US6051413-A.

XX

PD 18-APR-2000.

XX

PF 24-SEP-1998; 98US-0159539.

XX

PR 18-APR-1996; 96GB-0007791.

PR 17-OCT-1997; 97US-0953492.

PR 18-APR-1997; 97US-0844064.

XX

PA (SMK) SMITHKLINE BEECHAM CORP.

XX

PI Brown JR, Lawlor EJ, Wang M, Jaworski DJ;

XX

DR WPI: 2000-338311/29.

XX

DR N-PSDB; AAA14369.

XX

PT Novel vals polypeptides of valyl tRNA synthetase family useful for

PT treating otitis media, conjunctivitis, pneumonia and bacteremia

XX

XX comprises a specified amino acid sequence -

PS Claim 5; Columns 9-10; 21pp; English.

XX

CC This sequence represents Streptococcus pneumoniae valyl tRNA synthetase

CC (vals, NCIMB Deposit No. 40794). Streptococci are the cause of several

CC types of human diseases, including otitis media, conjunctivitis,

CC pneumonia, bacteremia, sinusitis, pleural empyema, endocarditis and

CC especially meningitis. The frequency of S. pneumoniae infections has

CC risen dramatically over the past 20 years, probably due to the emergence

CC of multiply antibiotic resistant strains and an increasing population of

CC immunocompromised people. Vals represents a target for new antibacterial

CC agents. Inhibition of tRNA synthetases such as vals leads to a reduction

CC in the levels of charged tRNA, which triggers a cascade of responses

CC (known as the stringent response) resulting in a state of dormancy in the

CC bacterium. Vals, its variants and fragments, anti-vals antibodies, vals

CC inhibitors and nucleotides encoding vals may be used in the diagnosis,

CC prevention and treatment of bacterial infections such as meningitis.

CC Vals can be used to screen compounds for inhibitory activity. Vals

CC may also be useful as an antigen for vaccination of a host to produce

CC specific antibodies which protect against bacterial invasion into

CC damaged tissues. Such antibodies could, for example, prevent the

CC adherence of bacteria to wounds. Nucleotides encoding vals may be used as

CC diagnostic reagents and therapeutic or prophylactic agents, particularly

CC for genetic immunisation.

XX

SO Sequence 883 AA;

Query Match 3.5%; Score 130; DB 21; Length 883;

Best Local Similarity 20.4%; Pred. No. 0.046;

Matches 104; Conservative 63; Mismatches 168; Indels 176; Gaps 24;

QY 192 DVF-----KQRYVFAFPSSVGKFGDPTDMMMPRHRTGDSVRRVYAGADNRAEYS 243

DB 26 DVFPGSGDOKAKPYSTVIPPNTGKL-----HLG-----HA 57

QY 244 KDNKPYKPYFAAVSMQYKADYAMTIGPGSTDRYL-TSMGVEDIREENNPRIEVRG 302

DB 58 WDT-TLQDIIIRKRMQGF-----TLWLPQMDHAGIATQAKYERLRKRGISRIIDLGR 110

QY 303 ---IKQIWKEMASADOATRIRYA-SKYAQASANYKNSIGMNRGLARLDVIGRKRAEER 357

DB 111 ESFLTKVMEWKD---EYATTIKEQMGKMGLSVDYSRERPTLDEGLS-----KAVRK 158

QY 358 AFDADWIRK-----NGKSAVIGD-----VLSLEAYIKEGAKANEMTYL----- 396

DB 159 VFVDLYRKGMVYRGEFTIMDPAARTALSDIEVIHKDVEGAFYHMYMLEDGSRVLEVAT 218

QY 397 --SETLFGTGEVVRFAOFANALATNPDAHGIKLSDDKYDYLPS-----LDRKYLPAM 449

DB 219 TRETFMGDV-----AVAVNPE-----DPRYKDLIGNVILPIANKLPIY 259

QY 450 LD-----IVRRIRPADKLPIDFKNVIDKKFGDTKKYADVFDKSVVPYSDK 496

DB 260 GDEHADPEFGTGVYKITPAHPDNDFLVQGRHNPQYVVMNDGTMDLAFEEFGM---DR 316

QY 497 FHAMLSMDEKFAKALENDPAVELSKSVIAARA-----IQAD 535

DB 317 FEAR-----RAVAKLEEIGALVKIEKRVASHGSEBTGVVPERLSTQFMKMDLAKN 371

QY 536 AMANAYAIKGRKRLFPAGLREMYGRALPS-----DANPTMRMSYGSIKGYEPQDG 586

DB 372 AIANODEDK-----VEFYPPRFNDFTLQWMEVHDWVISRLQWNGH-----QIP 416

QY 587 AMYNHTTGKGVLEKODPKSDEFAVOENIID 617

DB 417 AMYN--ADGEMVYGEBAPEGDGTODEDYLD 445

RESULT 4

AAW21725

ID AAW21725 standard; protein; 1194 AA.

XX

AC AAW21725;

XX

DT 01-OCT-1997 (first entry)

XX

DE Modified streptokinase, rSK6mut.

XX

KM Plasminogen-binding fragment; streptokinase; degradation; MBP;

KW thrombolytic agent; blood clot; bolus; maltose-binding protein.

XX

OS Streptococcus equisimilis.

XX

XX Synthetic.

XX

XX Key

XX Location/Qualifiers

FT Protein 1..381

FT /label= Maltose-binding_protein

FT /note= "acts as blocking group"

FT Protein 382..1194

FT /label= Modified_streptokinase

FT Misc-difference 391

FT /label= R10A

FT Misc-difference 417

FT /label= R36A

FT Misc-difference 426

FT /label= R45A

FT Misc-difference 432

FT /label= R51A

FT Misc-difference 440

FT /label= R59A

FT Misc-difference 766

FT /label= K385A

FT /note= "Given in the specification as K386A"

XX PN WO9641883-A1.
 XX PD 27-DEC-1996.
 XX PF 07-JUN-1996; 96WO-US09640.
 XX PR 09-JUN-1995; 95US-0488940.
 XX PA (HARD) HARVARD COLLEGE.
 XX PL Reed GL;
 XX DR WPI: 1997-065469/06.
 XX PT Modified forms of streptokinase resistant to enzymatic cleavage -
 PT useful as thrombolytic agents in treating thrombosis and in medical
 PT equipment
 XX PS Claim 16; Page 44-48; 65pp; English.
 XX CC This sequence represents a fusion protein between maltose-binding
 CC protein and a modified form of the plasminogen-binding fragment of
 CC streptokinase containing 6 point mutations. This modified streptokinase
 CC has an in vitro degradation rate at least 2 times slower than that of
 CC native streptokinase. Compounds containing modified streptokinases
 CC are specifically used as thrombolytic agents for dissolving blood
 CC clots in vivo in a mammal, preferably at a dose of 20000 U/kg,
 CC optionally as a bolus rather than by continuous infusion.
 XX SQ Sequence 1194 AA;

Query Match 3.4%; Score 127.5; DB 18; Length 1194;
 Best Local Similarity 20.1%; Pred. No. 0.12; Mismatches 263; Indels 167; Gaps 35;
 Matches 133; Conservative 99;

QY 75 GIVSDQGL-----IFTNHCYGAIQSQSTVDHDLRDGFVSRMGEELEPIP--GLS 125
 DB 571 GDTISQELLAQAQSIILKNKHPGYTIERDSSI--VTHNDIFRTI--LPMDEFTYR 624
 QY 126 VKYLKIKYVNDKVGOLKGTIDEMERLKAQVCOELAKKE----- 167
 DB 635 VKNRQDAVRINKR-----SGLNEEINNTDLISEKYYVLKGGKPYDPDRSHLKLFTIKY 679
 QY 168 -NADENQCIYEPFYSNNY--FLIVYDFEKDVRMVFAPSSVCKFG-----GDTDN 216
 DB 680 VDVQNELLSQQLLTASERNLDYDPDKAKLY--NNLDAFGIMDTLTGXVED 736
 QY 217 WKMPRHGDSFVFRVYAGADNRPAEYSKDNKPKPVYFAVSMOGYKADYAMTIGPGS 276
 DB 737 ---NHDDTNRIITYVMG--KRP---EGENASYH-----LAYDADRYEEEREYVS 778
 QY 277 TDRYLTMSGVEIRINENNPR-----EVRGKIGIKEMASDAQARIKASYAQC-- 328
 DB 779 YLRV-TGPIPIPNPDKNNSQLVSVAGTVEGTNODISLKEFEIDLTSRPAIGGTEGGL 837
 QY 329 ---SANYKNSIGMNGRLARLDVIGRKAERAFADWIRKN-----GKSAYYGVYL 376
 DB 838 SPSKSPATDSGAMGHKLEKADLL--KAIQEQLIAN-VHSNDYFEVIDAFADATITBRN 894
 QY 377 SLEKAYVEGAKANREMYTLSETLFGGTEVVRFAQFANALATNPAAHAGILKSLDDKY-- 434
 DB 895 GKVYFADKDG--VLTPTPOVQEFLLSGHVRVRYKE-----KPTIONA--KSVDEYTV 944
 QY 435 -----KDYLPPL--DRKVLPLAMLDIVRRIPADKLPDIFKAVIDKKFGDTKRYADPV 485
 DB 945 OPTPLNPDDEFPGKLKTLTL--AIGDITTSQELLAQAQSIILKNKHPG---YTYIE 998
 QY 486 FDKSVVPYSDEKHAMIKSNDKEKFAKAEKDPAVELSKSVIAAARAQADAMANA--YAE 544
 DB 999 RUSSIVTHNDIFRTILPMDQFTYRVKNRQAYRIKKSGSNEEINNTDLISEKYYVLK 1058
 QY 545 KKK-----RLFFAGLREMTPGRALPSDANFT---MMMSGISIKGYEPDGA--W 588

DB 1059 KGEKPYDPEDRSHLKLTFTIKYVDVDTNELLSKSEQLLTASERNLDYDPDKAKLY 1116
 QY 589 YN-----YHTGKGVLEKQDPKSEFAVQENIILFR--TFKNYGR--YAENGQJHI 635
 DB 1117 YNNLDAFGIMDTYLTGK-----VEDNHDDTNRIITYVMGKRPEGENASYHL 1162
 QY 636 AF 637
 DB 1163 AY 1164
 RESULT 5
 ID AU38091
 ID AU38091 standard; Protein: 883 AA.
 AC AU38091;
 XX 14-FEB-2002 (first entry)
 DE Streptococcus pneumoniae cellular proliferation protein #520.
 KW Antisense: prokaryotic cellular proliferation protein;
 KW antibiotic: antibacterial; drug design.
 OS Streptococcus pneumoniae.
 XX WO200170955-A2.
 XX 27-SEP-2001.
 PD 21-MAR-2001; 2001WO-US09180.
 PF 21-MAR-2000; 2000US-191078P.
 XX 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 PI WPI: 2001-611495/70.
 DR N-PSDB; AAS55950.
 DR New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 PT
 XX Example 3; Seq ID NO 13684; 51pp; English.
 XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Streptococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.


```
RESULT 7
AAM21723
ID AAM21723 standard; protein: 800 AA.
XX
AC AAM21723;
XX
DT 01-OCT-1997 (first entry)
XX
DE Plasminogen-binding domain from streptokinase.
XX
KM Plasminogen-binding fragment; streptokinase: degradation;
XX thrombolytic agent; blood clot; bolus.
XX
OS Streptococcus equisimilis.
XX
PN W09641883-A1.
XX
PD 27-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US09640.
XX
PR 09-JUN-1995; 95US-0488940.
XX
PA (HARD ) HARVARD COLLEGE.
XX
PI Reed GL;
XX
DR WPI; 1997-065469/06.
XX
PT Modified forms of streptokinase resistant to enzymatic cleavage -
PT useful as thrombolytic agents in treating thrombosis and in medical
PS equipment
PS
XX Claim 2; Page 35-37; 65pp; English.
CC This sequence represents the plasminogen-binding fragment of
CC streptokinase which lacks the N-terminal 14 amino acids. This
CC modified streptokinase has an in vitro degradation rate at least
CC 2 times slower than that of native streptokinase. Compounds
CC containing modified streptokinases are specifically used as
CC thrombolytic agents for dissolving blood clots in vivo in a
CC mammal, preferably at a dose of 20000 U/kg, opt. as a bolus
CC rather than by continuous infusion.
XX
SQ Sequence 800 AA;
Query Match 3.3%; Score 122.5; DB 18; Length 800;
Best Local Similarity 19.3%; Pred. No. 0.18;
Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;
OY 75 GIVSDQGL-----IFTNHCYGAIQSQTVDHDLRDFGVSRMGEELEPIP---GLS 125
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 177 GDTITSQELLAQAOSILNKNHPGTYIERDSSI---VTHDNDIEFRTI---LPMDEFTYR 230
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 126 VVYLKIVAVTQKVEGQLKGITDEMERLKAQEVCOELAKKE----- 167
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 231 VKNRQAVRIINK-----SGLNEEINMTDLISEKYYVLKKGKPPYDPDRSHLKTITKY 285
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 168 -NADENOLCIVEPFYSNNEX---FLIVYDFKQVRVFAFPSPSVGKFG-----GDTDN 216
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 286 VDVDTNELLSQLLTASERNLDFRDLIDPRDKAKILY---NNLDAFGIMDYTLTGKVED 342
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 217 WMMPRHTGDFSVFRYAGADNRPAEYSKDNKPYKPYFAVAVSMOGYKADYAMTIGFPGS 276
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 343 ----NHDDNRIITYMG--KRP---EGENASYHLAY-----D 371
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 277 TDRY-----LTSWGEVDRIENENNPTI-----EVRGIGKIWKREMSADQAT 318
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 372 KDRYTEEREVREYSLRYTGTPIPDNPNDRKNNSQLVYSVAGTEGVNODISLKFEEIDLTS 431
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 319 RIKYASKYQ-----SANTYKNSIGMNRGLARLDVIGKRAEERAFADWIRKN----- 366
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 432 RPAHGKTEQGLSPKSKPEATSDSGAMSHLEKADLL--KAIOEQLIAN--VHSNDYFEVI 488
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```

OY 367 --GRSAVYGDVLSLEKAYKEGAKANREMYLEFTLFGTEVYRPAQFANALATNPDAH 424
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 489 DEASDATITDRGKRYFFADKDS-VTLPEQYQGEFLLSGHVAVRKE-----KPIQNO 540
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 425 GILKSLDKY-----KDYLPSL-DRKYLPAMLDIVRRRIADKLPDIEKNVIDK 473
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 541 A--KSDVEYTVQFTPLNDDDFRPGDKDTKLKTL--AIGDTITSQELLAQAOSILKN 596
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 474 FKGDTRKYNADYFDDKSVYPYSDKFHAMLKSMCKEFKAKAIEPDPAVELSKSYIAARAIO 533
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 597 HPG---YTIYERDSSIVYTHDNDIFRTILPMQEFTYRKKNQAYRINKSGLEENINN 652
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 534 ADAMANA-YAIEKGR-----RLFFAGLRMYPGRALPSDANFT---MRMSYCSI 578
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 653 TDLISEKYYVLKKGKPPYDPDRSHLKLFTIKYVDVDTNELLSQQLLTASERNLDFRDL 712
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 579 KGEPODGA--WYN-----YHTGKGVLEKODPKSDEFAVOENILDLFR--TKNYG 625
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 713 --YDPRDKAKLLYNMLDAFGIMDYTLTGK-----VEDNHDDNRIITYMG 756
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
OY 626 R--YAENGQLHTAF 637
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 757 KRPEGENASYHLAY 770
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 8
AAM21728
ID AAM21728 standard; protein: 813 AA.
XX
AC AAM21728;
XX
DT 01-OCT-1997 (first entry)
XX
DE Wild type plasminogen-binding fragment of Streptokinase.
XX
KM Plasminogen-binding fragment; streptokinase: degradation; MBP;
XX thrombolytic agent; blood clot; bolus; maltose-binding protein.
XX
OS Streptococcus equisimilis.
XX
PN W09641883-A1.
XX
PD 27-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US09640.
XX
PR 09-JUN-1995; 95US-0488940.
XX
PA (HARD ) HARVARD COLLEGE.
XX
PI Reed GL;
XX
DR WPI; 1997-065469/06.
XX
PT Modified forms of streptokinase resistant to enzymatic cleavage -
PT useful as thrombolytic agents in treating thrombosis and in medical
PS equipment
PS
XX Example 1; Page 12-13; 65pp; English.
CC This sequence represents the wild type plasminogen-binding fragment
CC of streptokinase. This fragment was used in the design of a
CC modified streptokinase has an in vitro degradation rate at least
CC 2 times slower than that of native streptokinase. Compounds
CC containing modified streptokinases are specifically used as
CC thrombolytic agents for dissolving blood clots in vivo in a
CC mammal, preferably at a dose of 20000 U/kg, opt. as a bolus
CC rather than by continuous infusion.
XX
SQ Sequence 813 AA;
Query Match 3.3%; Score 122.5; DB 18; Length 813;
```


QY 534 ADAMANA-YAIEKGR-----RLFFAGLRKMPGRALPSPDANFT---MMSYSGSI 578
 Db 1034 TDLISEKYYVLKKGKRPDPDRSHLKLFTIKYVDVDTNELLSQQLLTASERNLDFRDL 1093
 QY 579 KGYEPQDGA--WYN-----YHTTGKGVLEKODPKSDFAVOENILDLFR--TKNYG 625
 Db 1094 --YDPRDKAKLLYNLNLDAFGIMDYTLTGK-----VEDNHDDTNRIITYVMG 1137
 QY 626 R--YAENGQHLIAF 637
 Db 1138 KRPEGENASYHLAY 1151
 RESULT 10
 AAW21724
 ID AAW21724 standard; protein; 1194 AA.
 AC AAW21724;
 XX
 XX 01-OCT-1997 (first entry)
 DE Modified streptokinase, rSK5mut.
 XX
 KM plasminogen-binding fragment; streptokinase; degradation; MBP;
 KM thrombolytic agent; blood clot; bolus; maltose-binding protein.
 XX
 OS Streptococcus equisimilis.
 XX
 FH key Location/Qualifiers
 FT Protein 1..381
 FT /label= Maltose-binding-protein
 FT /note= "acts as blocking group"
 FT Protein 382..1194
 FT /label= Modified-streptokinase
 FT Misc-difference 391
 FT /label= R10A
 FT Misc-difference 417
 FT /label= R36A
 FT Misc-difference 426
 FT /label= R45A
 FT Misc-difference 432
 FT /label= R51A
 FT Misc-difference 440
 FT /label= R59A
 XX
 PN MO641883-A1.
 XX
 PD 27-DEC-1996.
 XX
 PF 07-JUN-1996; 96MO-US09640.
 XX
 PR 09-JUN-1995; 95US-0488940.
 PA (HARD) HARVARD COLLEGE.
 PI Reed GL;
 XX
 DR WPI; 1997-065469/06.
 XX
 PT Modified forms of streptokinase resistant to enzymatic cleavage -
 PT useful as thrombolytic agents in treating thrombosis and in medical
 PT equipment
 XX
 PS Claim 15; Page 41-44; 65pp; English.
 XX
 CC This sequence represents a fusion protein between maltose-binding
 CC protein and a modified form of the plasminogen-binding fragment of
 CC streptokinase containing 5 point mutations. This modified streptokinase
 CC has an in vitro degradation rate at least 2 times slower than that of
 CC native streptokinase. Compounds containing modified streptokinases
 CC are specifically used as thrombolytic agents for dissolving blood

CC clots in vivo in a mammal, preferably at a dose of 20000 U/kg,
 CC optionally as a bolus rather than by continuous infusion.
 XX
 SO Sequence 1194 AA:
 Query Match 3.38; Score 122.5; DB 18; Length 1194;
 Best Local Similarity 19.38; Pred. No. 0.33;
 Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;
 QY 75 GITYSDGL-----IFNHHGCGYGAIOSQSTVHDHDIROGFSVRTMGEELPIP---GLS 125
 Db 571 GDTTSGELLAQAOSILNKNNHPGYTIYERDSSI--VTHMDIRFTI---LPMQEFYR 624
 QY 126 VKYLKIKIVYTDKVEGOLKGITDEMERLKAQVEQELAKKE----- 167
 Db 625 VKNEQAVRIKK-----SGINEEINNTDLISEKYYVLKKGKRPDPDRSHLKLFTIKY 679
 QY 168 -MADENOLCIVEPPYSNNY--FLIYDVKKDYRMVFPAPSSVKGFG-----GDTDN 216
 Db 680 VDVPDTNELKSEQLLTASERNLDFRDLYPDPRDKAKLLY---NNLDARGIMDYTLTGKVED 736
 QY 217 MMWRHHTGDFSVFRVYAGADNRPAPAEYSKDNKPYKPYFAAASMGYKADVAMTIGPPGS 276
 Db 737 ---NHDDTNRIITYVMG--KR--EGENASYHLAY-----D 765
 QY 277 TDRY-----LTSWGEDRIENENPRI-----EVRSIKGIMKRAMSADQAT 318
 Db 766 KDRYTEEREVEYSLRYTGPIPNPNNDKNSQOLVSVAGVEGTNDISLKFIEDLTS 825
 QY 319 RIKYASKYAO-----SANYKNSIGMNGRLADYIGKRAERAPADWIRKN----- 366
 Db 826 RPAHGGTEQGLSPKSKRFATDSGAMSHKLEKADLL--KAIOEOLIAN-VHSNDYFEVI 882
 QY 367 --GKSAYGVDLSSLEKAYKEGAKANEMTYLSTLGGTFVFAQFANALANPDAAH 424
 Db 883 DFASDATTDRNGKVPADKDG--VLTPOVQEFLLSGHVRVRYKE-----KPIQNO 934
 QY 425 GILKSLDDKY-----KDYLPSTL-DRKVLPAMLDIVRRIRPADLPDIFKNVIDKK 473
 Db 935 A--KSVDEVEYVQFTPLNPDDDFRGLKDTIKLTLT--AIGDITTSQELLAQAOSILNK 990
 QY 474 FKGDTKKYADEFVDPKSVVPSDKFHAMKSMKDKKFAKALEKDPAVELSKSVIAARAIO 533
 Db 991 HPG--YTIYERDSSIYTHNDIFRTILPMDQEFYTVKVRQEAQVIRKNSGNEIRIN 1046
 QY 534 ADAMANA-YAIEKGR-----RLFFAGLRKMPGRALPSPDANFT---MMSYSGSI 578
 Db 1047 TDLISEKYYVLKKGKRPDPDRSHLKLFTIKYVDVDTNELLSQQLLTASERNLDFRDL 1106
 QY 579 KGYEPQDGA--WYN-----YHTTGKGVLEKODPKSDFAVOENILDLFR--TKNYG 625
 Db 1107 --YDPRDKAKLLYNLNLDAFGIMDYTLTGK-----VEDNHDDTNRIITYVMG 1150
 QY 626 R--YAENGQHLIAF 637
 Db 1151 KRPEGENASYHLAY 1164
 RESULT 11
 AAW21726
 ID AAW21726 standard; protein; 1194 AA.
 AC AAW21726;
 XX
 XX 01-OCT-1997 (first entry)
 DE Streptokinase/maltose binding protein fusion protein, rSK.
 XX
 KM plasminogen-binding fragment; streptokinase; degradation; MBP;
 KM thrombolytic agent; blood clot; bolus; maltose-binding protein.
 XX
 OS Streptococcus equisimilis.
 XX


```

FH Key Location/Qualifiers
FT Protein 1..381
FT /label= Maltose_binding_protein
FT /note= "acts as blocking group"
FT 382..1194
FT Protein /label= Streptokinase
XX
XX MO9641883-A1.
XX
XX 27-DEC-1996.
XX
XX 07-JUN-1996; 96MO-US09640.
XX
XX 09-JUN-1995; 95US-0488940.
XX
XX (HARD ) HARVARD COLLEGE.
XX
XX Reed GL;
XX
XX WPI; 1997-065469/06.
XX
XX Modified forms of streptokinase resistant to enzymatic cleavage -
XX useful as thrombolytic agents in treating thrombosis and in medical
XX equipment
XX
XX Example 1; Page 11-12; 65pp; English.
XX
XX This sequence represents a fusion protein between maltose-binding
XX protein and the plasminogen-binding fragment of streptokinase.
XX This fusion protein was used in the design of modified streptokinase
XX has an in vitro degradation rate at least 2 times slower than that of
XX native streptokinase. Compounds containing modified streptokinases
XX are specifically used as thrombolytic agents for dissolving blood
XX clots in vivo in a mammal, preferably at a dose of 2000 U/kg,
XX optionally as a bolus rather than by continuous infusion.
XX
XX Sequence 1194 AA:
XX
XX Query Match 3.3%; Score 122.5; DB 18; Length 1194;
XX Best Local Similarity 19.3%; Pred. No. 0.33;
XX Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;
XX
XX 75 GIVNSDGL-----IFNTHCGGATIOSQSTVDHDLRDPFSRTMGCELPIP---GLS 125
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 571 GDITTSSELLAQASILINKHHPGTTIERDSSI---VTHNDIFRIT---LPMOQETFR 624
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 126 VKYLRKIVKTVDRVEGLKGITDEMERLRKAQVCEOLARKE----- 167
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 625 VKNREQAYRINKK-----SGLINEINNTDLISEKYVYLKKGKPYDPDRSHLKLFTIX 679
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 168 -MADENOLCIVEPFYSNNF---FLIYDYFKDYRMVFAFPSSVGRKG-----GPTDN 216
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 680 VVDVTNELKSEQLLTASEKNLDFRDLYPDRAKLILY---NNLDARGINDYTLTGKVED 736
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 217 WMPMRHTGDSVFVRVYAGARNRPAEYSKDNKPKYKPYFAVSMQGYKADYAMTIGPPGS 276
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 737 -----NHDDTKRITITVYG--KRP---EGENASTHLAY-----D 765
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 277 TDRT-----LTSKGVEDRIENENNRI-----EVRCIKGIGIRKAMSADQAT 318
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 766 KDRYTEEREVEVSYLRTGTPIPDNPDKNNNSOLVSVAGTVEGTNDISLKFEIDILTS 825
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 319 RITYASKYVAG-----SANYNKNSIGMNRGLARDVIGRKAEEAFADWIRKN----- 366
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 826 RPAHGGATEOGLSPKSKPFTDSGAMSHKLEKADL--KAIQEQLAN-VHSNDYEVI 882
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 367 --GKSAYGVGLVSLSEKAYEGAKANREMYLSETTLFGTEVEVFAFANALATPNDAHA 424
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 883 DPASDAITITBRNGKVYRADDS--VTLPIQPVQEFLLSGHVRVYKKE-----KPIQNG 934
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 425 GILKSLDDKY-----KDYLPST--DRKVLPAMLDIYRRRIADKLDPIDFNKVIDKK 473
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 935 A--KSDVVEYTVQFTPLNPDDEFRLPKDKYRLKLT--AIGDITTSQELLAQASILINK 990

```

```

XX 474 FKGDTRKYADPFVDRKSVYPSDKFHMLKSMDEKPAKALEKDPVAVLSKVIAARAIO 533
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 991 HPG-----YTIYERDSSIVTHDNDIFRTILPMDDEFYRVNRQAYINKSGLINEINN 1046
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 534 ADMAANA-VAIEKGR-----RLFPAGLRMPGRALPSDANFT--WRMSYGS 578
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1047 TDLISEKYIVLKGEKRYDPDRSHLKLFTIKYVDVTNELLSQQLTASERNLDFRDL 1106
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 579 KGYEPQDGA--WYN-----YHTTGKGVLEKODPKSDFAVQENILDLFR--TKNYG 625
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1107 --YDPRDKAKLTVNNMDAFGIMDYTLTGK-----VEDNHDTNRIITVYMG 1150
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 626 R--YAENGQLHAF 637
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 1151 KRPEGENASYHLAY 1164
XX
XX RESULT 12
XX AAY34536
XX ID AAY34536 standard; Protein; 951 AA.
XX
XX AAY34536;
XX
XX 25-AUG-1999 (first entry)
XX
XX Porphyromonas gingivalis protein PG67.
XX
XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX vaccine; antigenic.
XX
XX Porphyromonas gingivalis.
XX
XX MO929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998; 98MO-AU01023.
XX
XX 04-AUG-1998; 98AU-0005028.
XX
XX 10-DEC-1997; 97AU-0000839.
XX
XX 31-DEC-1997; 97AU-0001182.
XX
XX 30-JAN-1998; 98AU-0001546.
XX
XX 10-MAR-1998; 98AU-0002264.
XX
XX 09-APR-1998; 98AU-0002911.
XX
XX 23-APR-1998; 98AU-0003128.
XX
XX 05-MAY-1998; 98AU-0003338.
XX
XX 22-MAY-1998; 98AU-0003654.
XX
XX 29-JUL-1998; 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margolis MB, Patterson MA;
XX Ross BC, Rothel LJ, Webb EA;
XX
XX WPI; 1999-385613/32.
XX
XX N-PSDB; AAX91754.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
XX gingivitis
XX
XX Claim 1; Page 526-527; 588pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX activity with a vaccine mechanism of action. The PG polypeptides can
XX be used as vaccines especially against Porphyromonas gingivalis. Probes can
XX be used to detect Porphyromonas gingivalis in standard hybridisation
XX assays. Porphyromonas gingivalis is involved in periodontal disease
XX especially gingivitis.

```

SO	Sequence	951 AA;	3.2%;	Score 117.5;	DB 20;	Length 951;
Query Match	Best Local Similarity	20.3%;	Pred. No. 0.63;			
Matches	95;	Conservative	65;	Mismatches	190;	Indels 119; Gaps 22;
QY	248	PKKPYFAVAVSOGGKADYAMTIGFPGSTIDYLSWG---- <td>302</td> <td></td> <td></td> <td></td>	302			
DB	263	PSNKRYSGLIMPTYGEDN--RYGF-----YLNNGGYFAFSYID-----LALRGE	307			
QY	303	-IKQIKWEAMSADQATRIKYSKKAOSANYWKNKSIGNMRGLARLDVIGKRRAERAFAD	361			
DB	308	IFSKKSWGISQASOKKKRYKNGSF--EANY-----	336			
QY	362	WIRKNGKSAVYGDVLS---SLEKAYKKGAKANREMYLSETLFGTEVYVFAQAMALA	417			
DB	337	LVSKSGDYVPEDYSKTSLNIRWHSODPKANP-----LQTLSANVNFATGSYFONSIN	391			
QY	418	TNPDAHA--GILKSLDDKYYKDLPSLDRKYLPLAMDI---VARRLRPADKLPDIFKNVIDK	472			
DB	392	TTYDVNATPATTTSSAVSYSRRKFPPTPSITGSM--DISQNMDDTVSLPLMLSLMSTR	450			
QY	473	-KFKGDTKKYADFVDSVVPYSDKFFHMLKMSDKKEFAKAEKPAVELSKSVIAAARA	531			
DB	451	YPKFKTKTVGPWRWYEXLSVGSGLRNSILTKEDDLQSLNLYVRPMKGMKMRSHVPSISLV	510			
QY	532	IQADAMANAVALIEKGRLLFFAGLRMY--PGRALPSDANFTMRNSYGSINKYEPQDGAW	588			
DB	511	PLDDYINILTMGVNVMWYTKGIRKSMWEDKKTFLPSDPTTYKFRRLY---DYSLSAG--	564			
QY	589	YNYHTTGAVYLEKODPKSDEFVQENILD-----FRTKNYGRVA-----	628			
DB	565	--LSTTLTGMEFPMWRP---FSFGGILMIRHRTPTVSFSYMPDFTKRYKGFWEELLEHT	618			
QY	629	-ENGOLHIAFLSN--NDITG-----GNSGSPYDKNGLRILIGLAFQDNWEA	670			
DB	619	DONGRLHTLLISPYEFQILFAPSMGNAGS-----VNSFSNNLEA	658			
RESULT 13						
AY34403						
ID	AY344403 standard; Protein: 953 AA.					
AC	AAV34403;					
XX						
DT	25-AUG-1999 (first entry)					
XX						
DE	Porphyromonas gingivallis protein PG67.					
XX						
KW	Porphyromonas gingivallis; PG; periodontal disease; gingivitis;					
KW	vaccine; antigenic.					
XX						
OS	Porphyromonas gingivallis.					
XX						
PN	W09929670-A1.					
XX						
PD	17-JUN-1999.					
XX						
PE	10-DEC-1998; 98WO-AU01023.					
XX						
PR	04-AUG-1998; 98AU--0005028.					
PR	10-DEC-1997; 97AU--0000839.					
PR	31-DEC-1997; 97AU--0001182.					
PR	30-JAN-1998; 98AU--0001546.					
PR	10-MAR-1998; 98AU--0002264.					
PR	09-APR-1998; 98AU--0002911.					
PR	23-APR-1998; 98AU--0003128.					
PR	05-MAY-1998; 98AU--0003338.					
PR	22-MAY-1998; 98AU--0003654.					
PR	29-JUL-1998; 98AU--0004917.					
XX						
PA	(CSLC-) CSL LTD.					
XX						

Pt	Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA,
Pl	Ross BC, Rothe LJ, Webb EA;
Dx	WP1; 1999-385613/32.
DR	N-PSDB; AAX9162L.
Pn	Antigenic Porphyromonas gingivalis peptides for preventing gingivitis
Pt	
Pd	
Ps	Claim 1; Page 373-375; 588bp; English.
Pf	
Cc	AAX91536 to AAX91801 encode two hundred and sixty six antigenic Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to AAX34563. AAX91802 to AAX91899 represent PCR primers used in the CC isolation of the PG polypeptides. The PG polypeptides have antibacterial activity with a vaccine mechanism of action. The PG polypeptides can be used as vaccines especially against Porphyromonas gingivalis. Probes can be used to detect Porphyromonas gingivalis in standard hybridisation assays. Porphyromonas gingivalis is involved in periodontal disease especially gingivitis.
Cc	
Cc	
Sq	Sequence 953 AA;
Oy	Query Match 3.2%; Score 117.5; DB 20; Length 953; Best Local Similarity 20.3%; Pred. No. 0.63;
Matches	Matches 95; Conservative 65; Mismatches 190; Indels 119; Gaps 22
Db	248 PYKPYFAVAASQGRADYATMTIFPFGSTDYLTSMWG---VEDRIENNNPRIEVRG - 302 : : : : : : : : : : : : : : : 265 PFSNKYSIGILMPYTGEDN---RYGF-----YLNRGYFAPSDID-----LALRGE 309
Oy	303 -IKGIWKEAMSADOTRIKIYASKYAQSANVWNKSIGNMRGLARDLDVGKRRAERAFAD 361 : : : : : : : : : : : : : : : 310 IFSKGSGWSIASQSKYKKRYKNCSF--EANY----- 338
Oy	362 WIRKNGKSAVVDVLS----SLEKAYKBESAKANNRENTISETLFGTETVVRRQAPANALA 417 : : : : : : : : : : : : : : : 339 IVSKSGDKYPDSDYKTSLNTRTHSHODKPANP-----LOTLSANVNVPATGSYEQNSLN 393
Db	418 TNPDDHA--GIKLSLDDCKDYLPISLRKVLPAMLDI---VRRRIPADKLPIFKNVIDK 472 : : : : : : : : : : : : : : : 394 TTYYDNARAIATRSSAVSVSRKFPGSPITSISM-DISONMMDTIVSLPLNLSTMSTR 452
Oy	473 -KFKGDITKYADFVEDEKSVVPSDFHAMLKSMDEKFEKAIEKDPAVELSKSVIAAARA 531 : : : : : : : : : : : : : : : 453 YPFKRTTRGVPERMYEKLSVYGSQLRNLSILTKEKDLLSNLVDPWKMGHRHSVIPSTLV 512
Oy	532 IOADMANAAYAIEKKRLFFAGLRENY---PGRALPDSANFTMKMSYGSIKCYEPQDGAW 588 : : : : : : : : : : : : : : : 513 PLLDIYNITLMGYNVMWTYKGIKRSWMNDKTFPLPDITYKFRRLY---DYSLSAG-- 566
Db	589 VYHTTTGGVLEKOPKSDSEFAVOENIIDL-----FRTKNYGRA----- 628 : : : : : : : : : : : : : : : 567 --LSTLTIGMEFPKRP-----FSFGNLIMIKRRFLPTVSFSYMPDPTRYKRGFWELLEHT 620
Oy	629 -ENGOLHIAFILSN--NDITG---GNSGSPVEDKNGRLIGLAFDGWMEA 670 : : : : : : : : : : : : : : : 621 DQNGKLHLTLTPYEPEQJIGAPSWMGAS-----VFSEFNLLLEA 660
Db	
RESULT 14	
ID	AAR88649
AC	AAR88649 standard; Protein: 711 AA.
XX	
XX	AAR88649;
XX	
XX	02-SEP-1996 (first entry)
XX	
XX	Neisseria meningitidis BZ163 transferrin receptor Tbp2 subunit.
KX	Transferrin receptor; Tbp2 subunit; deletion mutant; vaccine:
KW	passive immunisation; Immunotherapy; IM2169; IM2394.
XX	

OS Neisseria meningitidis (strain B2163).
 XX Key Location/Qualifiers
 FH Protein 1..711 /label= Tbp2_subunit
 FT Peptide 1..20 /label= signal_peptide
 FT Domain 21..366 /label= Domain_1
 FT /note= "N-terminal domain"
 FT 367..564 /label= Domain_2
 FT /note= "hinge domain"
 FT 565..711 /label= Domain_3
 FT /note= "C-terminal domain"
 PN MO953049-A2.
 PD 07-DEC-1995.
 XX 30-MAY-1995; 95MO-FR00701.
 PF 31-MAY-1994; 94FR-0006594.
 XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS.
 PA (TRGE) TRANSGENE SA.
 XX Jacobs E, Legrain M, Mazarin V, Lissolo L;
 PI Millet MBJ;
 XX WPI: 1996-030562/03.
 DR N-PSDB; AAT11244.
 XX
 PT Polypeptide(s) for vaccination against Neisseria meningitidis group
 PT B - comprising deletion mutants of transferrin receptor Tbp2
 PT subunit
 XX
 PS Disclosure: Page 82-87; 114pp; French.
 XX
 CC The present sequence is that of the N.meningitidis strain B2163
 CC transferrin Tbp2 subunit. The Tbp2 polypeptide has three
 CC domains (see features table); deletion mutants in which at least
 CC one of the domains is partially or totally deleted are claimed,
 CC provided that the first and second domains are not simultaneously
 CC partially or totally deleted. The positions of the 3 domains in
 CC B2163 are defined by alignment with the IM2169 sequence. The
 CC deletion mutant polypeptides of the invention can generate an
 CC immune response against N.meningitidis.
 CC
 SO Sequence 711 AA:
 Query Match 3.1%; Score 114; DB 17; Length 711;
 Best Local Similarity 18.7%; Pred. No. 0.81;
 Matches 134; Conservative 97; Mismatches 264; Indels 220; Gaps 34;
 QY 80 DOGIFNHHGCG-ATOSQSTVDHYLRDGFYSRTMG--ELPRLPSKYLKIKYKVT 136
 DB 57 DQG-----GYGFAKRLKRRNRHPQAKEDKVELNDNMEETGLPSKPONLPERQSVI 108
 QY 137 DKVEGQLGITDEMERLRKAQEOELAK--KENADENOLCIPEFYSNNEYFLIYDV 194
 DB 109 DKYK-----TDGSGNITYTSPYLQSNHONGSTNSGANO-----PKNE---VKD-Y 149
 QY 195 KDVRAMFAPPSVVGKFGGDTDNMMMPRH--GDFSVFRVYAGADNRPAYSKDNKP-- 248
 DB 150 KNFKYIYVS-----GMFY-KHAESEREFSKIKFKSGDDGYIFYHGKDPSPQLP 195
 QY 249 -----YKPY-YFAAVSMOGYKADDMYMTIGFGSGTDRIYLSMG-----VEDRIENNNP 236
 DB 196 TSEKVIYKGVHVTDEKQKNDILLET--SKGOGGRIYSGFGSDGDETTSNFTDSNLND 233
 QY 297 RIEVRGIGKIGIMWEAMS-----ADQATRIKYSKY-----AQGAN----- 331

DB 254 KHEGYFTSLLEVDFPSKULGKILIRNNRYTNATNDKTYTQYSSDAQTGCRFGKAI 313
 QY 332 -----YMKNSIGMNRGL-----ARLDVIGKRAEERAPAD 361
 DB 314 ATDKPPTGGTKLHPFVSDSSLSGFGFPGKEELGREFLSDDKKAVVGSAAKTKDKT--- 370
 QY 362 WIRKNGKSAVYGDVLSLEKAYKEGAKANREMTYLS--ETLPGCTVVRFAQFANL--- 416
 DB 371 ---ENGAVASGGTDAASNGACGTSSENKSLTIVLDAVELKLDCKEYOKIDNFNSMAQLY 427
 QY 417 -----ATNPAAHAGILKSLDDKYK-DYLPISDRKRVLPAMDIVRRRIIPADK 461
 DB 428 VDCIMIPILPETSESGNQANOGTNGTAFTRKFDHTPESDKKDAQGTGNAQASNN 487
 QY 462 LPDIFKAVIDKKRGDTKKADVFEDKSVYPSYDKRHAMLSKDKKEKFAKIKDPAVEL 521
 DB 488 AGDT-----NGKTKTY-----EVEVCCSNLYLKGM/LTRKN----- 519
 QY 522 SKSVIAARAIOADAMANAIAIEGKRLFPAGLREMYPGRALPSDANFTWRMSYGSIKY 581
 DB 520 SKS---AMQGESSQADAKTEQYGSMFLQG--ERTDEKIRSEQNIYR----- 555
 QY 582 EPQDGAWYNY--HTTGKG-VLEKODPKSDPEFAVOENILFTKKNYGRYAENGOLHAF 637
 DB 566 ---GSWYGHIASTSMGNSADKKEGNRAEFTV-----NGEKKIKGTI----- 606
 QY 638 LSNNDITGNGSGSPVDPKNGRLIGLAFDGNWEMSGDIEREPLORTISVDIRY 692
 DB 607 -----TAENROEATFTIDKIEGNGFSGT--AKTAEIGFDLDOKNTTRPKAYI 653
 RESULT 15
 AAB43567
 ID AAB43567 standard; Protein; 708 AA.
 AC AAB43567;
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated protein sequence SEQ ID NO:1012.
 XX
 KW Human; cancer associated gene; cancer antigen; detection; cancer;
 KW diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
 KW antidiabetic; antihistaminic; antihypertensive; antirheumatic; antitumor;
 KW antineoplastic; antineoplastic; antineoplastic; antineoplastic; antineoplastic;
 KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
 KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
 KW immune disorder; hematopoietic cell disorder; autoimmune disorder;
 KW allergic reaction; graft versus host disease; organ rejection;
 KW haemostatic; thrombolytic; cardiovascular disorder; infection;
 KW neurological disease; drug screening.
 XX
 XX Homo sapiens.
 PN MO200055350-AL.
 PD 21-SEP-2000.
 PF 08-MAR-2000; 2000MO-US05882.
 PR 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Ruben SM;
 PI WPI: 2000-587533/55.
 DR N-PSDB; AAC77776.
 XX Novel isolated nucleic acids comprising sequences encoding peptides
 PT useful for treating or diagnosing e.g. cancer -
 XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 22:36:01 : Search time 24 Seconds
(without alignments)
724.626 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MCMKLKSLILGALLLGASG.....LFMIDKMGCCPRLLIQLKLI 712

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/3A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/3B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/4A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/4B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCPUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	3.5	883	2	US-08-953-492-2
2	127.5	3.4	1194	2	US-08-488-940-18
3	122.5	3.3	800	2	US-08-488-940-4
4	122.5	3.3	813	2	US-08-488-940-3
5	122.5	3.3	1181	2	US-08-488-940-2
6	122.5	3.3	1194	2	US-08-488-940-1
7	122.5	3.3	1194	2	US-08-488-940-17
8	122.5	3.3	1027	4	US-08-446-137B-2
9	109	2.9	1577	2	US-08-793-824-2
10	106	2.9	972	3	US-08-335-844A-23
11	103.5	2.8	1114	4	US-08-811-583-2
12	102.5	2.8	833	4	US-09-514-302-3
13	102.5	2.8	1938	4	US-09-514-302-2
14	101.5	2.7	747	1	US-07-854-596B-40
15	101	2.7	433	1	US-08-417-492-2
16	100.5	2.7	638	1	US-08-712-241-6
17	100.5	2.7	885	1	US-08-484-105-14
18	100.5	2.7	885	1	US-08-484-106-14
19	100	2.7	15281	2	US-08-471-119A-2
20	99.5	2.7	722	4	US-08-961-083-84
21	99	2.7	725	2	US-08-816-105A-1
22	99	2.7	908	2	US-08-249-380-2
23	98.5	2.6	637	1	US-08-712-241-5
24	98.5	2.6	637	1	US-08-026-143B-4
25	98.5	2.6	637	4	PCT-US92-10621-4
26	98.5	2.6	637	5	PCT-US94-02233-4
27	97.5	2.6	468	4	US-08-961-083-10

28	97.5	2.6	1430	3	US-09-008-172-2	Sequence 2, Appl
29	97.5	2.6	1430	4	US-09-210-361-6	Sequence 6, Appl
30	97	2.6	890	4	US-09-342-648-10	Sequence 10, Appl
31	96.5	2.6	654	1	US-08-487-890A-12	Sequence 12, Appl
32	96.5	2.6	654	2	US-08-478-435-12	Sequence 12, Appl
33	96.5	2.6	654	2	US-08-337-483-12	Sequence 12, Appl
34	96.5	2.6	654	2	US-08-478-373-12	Sequence 12, Appl
35	96.5	2.6	654	3	US-08-474-671-12	Sequence 12, Appl
36	96.5	2.6	654	3	US-08-483-577A-12	Sequence 12, Appl
37	96.5	2.6	654	4	US-08-897-438-12	Sequence 12, Appl
38	96.5	2.6	654	4	US-08-637-654-12	Sequence 12, Appl
39	96.5	2.6	1128	1	US-08-111-939-2	Sequence 2, Appl
40	96.5	2.6	1181	1	US-08-053-614-4	Sequence 4, Appl
41	96.5	2.6	1181	1	US-08-316-397B-4	Sequence 4, Appl
42	96.5	2.6	1181	2	US-09-034-306-4	Sequence 4, Appl
43	96.5	2.6	1181	4	US-09-259-437-4	Sequence 4, Appl
44	96.5	2.6	1181	5	PCT-US93-09782-4	Sequence 4, Appl
45	96	2.6	1007	4	US-08-961-083-216	Sequence 216, App

ALIGNMENTS

RESULT 1
US-08-953-492-2
Sequence 2, Application US/08953492
Patent No. 5849555
GENERAL INFORMATION:
APPLICANT: Brown, James
APPLICANT: Jaworski, Deborah
APPLICANT: Lawlor, Elizabeth
APPLICANT: Wang, Min
TITLE OF INVENTION: NOVEL VALS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/953,492
FILING DATE: 17-OCT-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/844,064
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607991.8
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmil, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31458-4/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 883 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-953-492-2

Query Match 3.5%; Score 130; DB 2; Length 883;
 Best Local Similarity 20.4%; Pred. No. 0.0042;
 Matches 104; Conservative 63; Mismatches 160; Indels 176; Gaps 24;

```

QY 192 DVF-----KDVNFVFPSSVCKEGGDTNNMMPRTGDFSVFRVYAGADNRPAYXS 243
      ||| : : : : :
DB 26 DVFKEPSGDOKARPSIVIPPPNVTGKL-----HLG-----HA 57
      ||| : : : : :
QY 244 KKNKPKYKPYFAAVSMOGKADYAMTIGFPGSTDRYL-TSMGVEDRIENENNPRIEVRG 302
      ||| : : : : :
DB 58 WPT-PLQDITIKKRMQGD-----TLMLPGMDHAGIATQAKVEERLKGISRDGLR 110
      ||| : : : : :
QY 303 ---IKQIIMKEMASDAQATRIKYA-SKYAOSANYKNSIGNRGLARLDVIGRKAER 357
      ||| : : : : :
DB 111 ESPLFRVMEKMD---EYATTIKEQGMKGLSVDSREFTIDEGLS-----KAVRK 158
      ||| : : : : :
QY 358 ARADWTRK---NGKSAYVGD-----VLSLEKAYKEGAKANREMYL----- 396
      ||| : : : : :
DB 159 VEVDLTKKGMIVRGEIIMMDPAARTALSDIEYIHKDVGAFYHNMVLEDEGSRYLEVAT 218
      ||| : : : : :
QY 397 ---SETLFGTEVVRFAQFANALATNPDAHAGILKSLDDKYYKXLYPS-----LDRKVLPM 449
      ||| : : : : :
DB 219 TRPRTFEGVY-----AAVNP-----DPYKDLIGKNVILPLANKLIPY 259
      ||| : : : : :
QY 450 LD-----IVRRRIPADKLPDIFKNVIDKKEFGDTKKYADFEVFKSVVPYSDK 496
      ||| : : : : :
DB 260 GDEHADPEFGTVYKITPAHDNDPLVQGRHMLPQVNVNNDGDTMIDLAFESGM---DR 316
      ||| : : : : :
QY 497 FPAMLKSMKEKFAKLEMDPAVELSKSYIAARA-----IQAD 535
      ||| : : : : :
DB 317 FEER-----KAVYAKLEEALVKLEKRVHSGHSRIGVVVEPRLSTQMEVYKMDQLAN 371
      ||| : : : : :
QY 536 AANAIAIEKGRLEFAGLREMYPGRALPS-----DANFTMMSYSIKGYPEQDG 586
      ||| : : : : :
DB 372 ALANDQTEK-----VEFYPRFNDTFLQMMENVAHDWVISRLAMGH-----QIP 416
      ||| : : : : :
QY 587 ANYNHTTKGVLEKODPKSDEFAVOENILD 617
      ||| : : : : :
DB 417 AMYN--ADGEMVYGEARPEGDGTODEVDLD 445
      ||| : : : : :

```

RESULT 2

US-08-488-940-18
 ; Sequence 18, Application US/08488940
 ; Patent No. 5854049
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Guy L.
 ; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P. C.
 ; STREET: 225 Franklin Street
 ; CITY: Boston
 ; STATE: MA
 ; COUNTRY: USA
 ; ZIP: 02110-2804
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/488,940
 ; FILING DATE: 09-JUN-1995
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fraser, Janis K.
 ; REGISTRATION NUMBER: 34,819
 ; REFERENCE/DOCKET NUMBER: 05433/009001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617/542-5070
 ; TELEFAX: 617/542-8906
 ; TELE: 200154

INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1194 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-488-940-18

Query Match 3.4%; Score 127.5; DB 2; Length 1194;
 Best Local Similarity 20.1%; Pred. No. 0.012;
 Matches 133; Conservative 99; Mismatches 263; Indels 167; Gaps 35;

```

QY 75 GITVDOGL-----IFTHHCGYGAIOSQSVVDHDLRDGVSRTMGELPIP---GLS 125
      ||| : : : : :
DB 571 GDTITSQELLAQOSILKNHFGYTIYERDSSI---VTHDNDIFRTI---LPMDEPIYR 624
      ||| : : : : :
QY 126 VYLRRIYKVTVDKVBGQLGITDEMERLKAQOEVCOELAKKE----- 167
      ||| : : : : :
DB 625 VKNRQAVRINK-----SGLNEEINNTDLISEKYVLKKGKPYDPDRSHLKLFTIKY 679
      ||| : : : : :
QY 168 -NADENQCIYEPFYSNNEX---FLIYDVFEKDVAMVFPSSVCKFG-----GDTDN 216
      ||| : : : : :
DB 680 VDVDTNELKSQLLTASERNLDFRDLDYPRDKALLY---NNLDAFGIMDTLTKGVED 736
      ||| : : : : :
QY 217 WMMPHRTGDFSVFRVYAGADNRPAYSKDNKPKYKPYFAAVSMOGYKADYAMTIGFPGS 276
      ||| : : : : :
DB 737 ---NHDTNRIITYMG--KRP---EGENASYH-----LAYDADRYEEREYYS 778
      ||| : : : : :
QY 277 TDRYLTSMGVEDRIENENNPRI-----EVKGICQIMKEMASDAQATRIKASYAQ-- 328
      ||| : : : : :
DB 779 YLRV-TGPIPIPNPNPDKNNSQLVSVAGTVEGTNDISLKFPEIDLTSRPAHGKTEQGL 837
      ||| : : : : :
QY 329 ---SANYMNSIGNRGLARLDVIGRKAERAFADWIRKN-----GKSAYGVYL 376
      ||| : : : : :
DB 838 SPKSPFATDSGAMSHKLEKADLL--KAIQEQILIN-VHSNDYREVLDPASDATITDRN 894
      ||| : : : : :
QY 377 SLEKAYREGAKANREMYLSETLFGTEVVRFAQFANALATNPDAHAGILKSLDKY-- 434
      ||| : : : : :
DB 895 GKVYFADKDG--VTLPTQVQGEFLLSGHVRYRKE-----KPIQONQ--KSVDEVETV 944
      ||| : : : : :
QY 435 -----KDYLPSL-DKRVLPAMLDIYRRIRIPAKLPDIFKNVIDKKEFGDTKKYAPFV 485
      ||| : : : : :
DB 945 QPTPLNPDDFPRGKDKLKLTL--AIGDITTSQELLAQOSILKNHFG--YTYE 998
      ||| : : : : :
QY 486 FPKSVVPYSDKFPAMLSMDKEKFAKLEKODPAVELSKSVIAARAIOADAMANA-YAIE 544
      ||| : : : : :
DB 999 RDSSTVTHNDIFRTILPMDQETRYKKNRQAVRINKKSGINEINNTDLISEKYYVLK 1058
      ||| : : : : :
QY 545 KGR-----RIFFAGLREMYPGRALPSDANFT--MMSYSGIKGYEPQDGA--W 588
      ||| : : : : :
DB 1059 KCEKPYDPDRSHLKLFTIKYVDVDTNELKSEQLLTASERNLDPRDL--YPRDKAKLL 1116
      ||| : : : : :
QY 589 YN-----YHTTGKGVLEKODPKSDEFAVOENILDLEF--TKNYGR--YAENGOLHI 635
      ||| : : : : :
DB 1117 YNNLDAFGIMDTLTKG-----VEDNHDTNRIITYMGKRPGENASYHL 1162
      ||| : : : : :
QY 636 AF 637
      ||| : : : : :
DB 1163 AT 1164
      ||| : : : : :

```

RESULT 3

US-08-488-940-4
 ; Sequence 4, Application US/08488940
 ; Patent No. 5854049
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Guy L.
 ; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P. C.
 ; STREET: 225 Franklin Street

```

? CITY: Boston
? STATE: MA
? COUNTRY: USA
? ZIP: 02110-2804
?
? COMPUTER READABLE FORM:
?
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/488,940
? FILING DATE: 09-JUN-1995
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: Fraser, Janis K.
? REGISTRATION NUMBER: 34,819
? REFERENCE/DOCKET NUMBER: 05433/009001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617/542-5070
? TELEFAX: 617/542-8906
? TELEX: 200154
?
? INFORMATION FOR SEQ. ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 800 amino acids
? TYPE: amino acid
? STRANDEDNESS: not relevant
? TOPOLOGY: linear
? MOLECULE TYPE: protein
?
? US-08-488-940-4

```

Query Match 3.3%; Score 122.5; DB 2; Length 800;

Best Local Similarity 19.3%; Pred. No. 0.018; Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

```

QY 75 GITVSDGL-----IFTNHCYGAIQSQSYVDHDLRGFVSRMGELPIP---GLS 125
DB 177 GDTTSEELLAQASILNKNHGPTIYERDSI---VTHNDIFRTI---LPMQDEFTYR 230
QY 126 VKYLRKIVKTVDRVEGLKGTIDEMERLKAQVCOELAKE-----167
DB 231 VKNREQAVRINKK---SGLNEINNTDLISEKYVYLKKGKRPYDFDRSHLKLFTIKY 285
QY 168 -NADENOLCIVEPEPYSNNEY---FLIYDYFKVIRWVFAFPSSVGKFG-----GDTDN 216
DB 286 VDVDTNELLKSEQLLTASERNLDPRDLYDPRDAKILY---NNIDAFGIMDYTLTGVED 342
QY 217 WMPFRHTGDSVFRVYAGADNRAPEYSKDNKPYKPYFAAVSMQGYKADYAMTIGPPGS 276
DB 343 ---NHDDTRITITVWG---KRP---EGENASYHLAY-----D 371
QY 277 TDRY-----LTSNGVEDRIENENNPR-----EVGIGIKGIMKEAMSADQAT 318
DB 372 KDRTEEREVEYSYLRGTGPIDNPNDKNNSQLVSVAGTVEGTNDISLKEFEIDLTS 431
QY 319 RIKYASKYAO-----SANYKNSIGMNRGLARLDVIGRRKRAEERAFADWIRKN-----366
DB 432 RPAHGKTEBGLSPKSKFPATDSGAMSHKLEKADL--KAIQEQLIAN--VHSNDYREVI 488
QY 367 --GKSAVYGVLSLEKAYEGAKANREMYLSETLFGTEVVVRFQAFANALATNPDAHA 424
DB 489 DFASDATITTRNGKVFYFADKGS--VTLPTQPVQEFLLSGHVRYRKE-----KPIQNG 540
QY 425 GILKSLLDDKY-----KDYLPST--DRKVLPRAMLDIYRRRIIPAKLDPIDFRNVIDK 473
DB 541 A--KSVAVETTVQFTPLNPDDEPRGKDKLKLKTL--AIGDITTSOELLAQASILNKN 596
QY 474 FKGDTRKYADFPEDKSVYPSDKFHAMLSMDKEKFAKATEKDPAAVLSKSVIAARAIO 533
DB 597 HPG---YTIYERDSSILVTHNDIFRTILPMDQFTYRVKNRREGAVYINNKSGINEIRNN 652
QY 534 ADAMANA-YAIEKCK-----RLFAGLRKEMTGGALPSDANFT--MRMSYGST 578
DB 653 TDLISEKYVYLKKGKRPYDFDRSHLKLFTIKYVDVTNELKSEQLLTASERNLDFRDL 712

```

```

QY 579 KGYEPOGA--WYN-----YHTGKGVLEKQDPKSDFAVOENILDFR--TKNYG 625
DB 713 --YDPRDAKALLYNNDLAFGIMDYTLTGK-----VEDNHDDTRITITVWG 756
QY 626 R--YAENGOLHIAF 637
DB 757 KREGENASYHLAY 770

```

```

RESULT 4
US-08-488-940-3
; Sequence 3, Application US/08488940
; Patent No. 5854049
; GENERAL INFORMATION:
; APPLICANT: Reed, Guy L.
; TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,940
; FILING DATE: 09-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 05433/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
;
; INFORMATION FOR SEQ. ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 813 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
? US-08-488-940-3

```

Query Match 3.3%; Score 122.5; DB 2; Length 813;

Best Local Similarity 19.3%; Pred. No. 0.019; Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

```

QY 75 GITVSDGL-----IFTNHCYGAIQSQSYVDHDLRGFVSRMGELPIP---GLS 125
DB 190 GDTTSEELLAQASILNKNHGPTIYERDSI---VTHNDIFRTI---LPMQDEFTYR 243
QY 126 VKYLRKIVKTVDRVEGLKGTIDEMERLKAQVCOELAKE-----167
DB 244 VKNREQAVRINKK---SGLNEINNTDLISEKYVYLKKGKRPYDFDRSHLKLFTIKY 298
QY 168 -NADENOLCIVEPEPYSNNEY---FLIYDYFKVIRWVFAFPSSVGKFG-----GDTDN 216
DB 299 VDVDTNELLKSEQLLTASERNLDPRDLYDPRDAKILY---NNIDAFGIMDYTLTGVED 355
QY 217 WMPFRHTGDSVFRVYAGADNRAPEYSKDNKPYKPYFAAVSMQGYKADYAMTIGPPGS 276
DB 356 ---NHDDTRITITVWG---KRP---EGENASYHLAY-----D 384
QY 277 TDRY-----LTSNGVEDRIENENNPR-----EVGIGIKGIMKEAMSADQAT 318

```


APPLICATION NUMBER: US/08/488,940
 FILING DATE: 09-JUN-1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: Fraser, Janis K.
 REGISTRATION NUMBER: 34,819
 REFERENCE/DOCKET NUMBER: 05433/009001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1194 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-488-940-1

Query Match 3.3%; Score 122.5; DB 2; Length 1194;
 Best Local Similarity 19.3%; Pred. No. 0.035;
 Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

QY 75 GITVSDQGL-----IFTHHCGYGAIGOSQTVVDHDLRDGFVSRMGELPIP---GLS 125
 DB 571 GDTITSQELLAQAOSILNKNHPGTYIERDSSI---VTHDNDIFRTI---LPMDOEFYR 624
 QY 126 VYLRKIVAVTDKVEGOLGIDEMERLKAQEVCOELAKKE----- 167
 DB 625 VKNRQAVYRINKK-----SGLNEEINNTDLISEKYYVLKKGEPYDPFRSHLKFTIY 679
 QY 168 -NADENQOLIVEPFYSNNY---FLIVYDFKDYRMVFPAPSSVGF-----GDTDN 216
 DB 680 VVDVNTLLKSQLTLASERNLDFRLYDPRDKAKLLY---NNLDAFGIMDYTLTGKVD 736
 QY 217 WMPVRHTGDSVFRVYAGADNRPAREYSKDNKPKPYFAAVSMQGYKADYAMTIGFPGS 276
 DB 737 ----NHDDTNRIITYVMG---KRP---EGENASYHLAY-----D 765
 QY 277 TDYR-----LTSWGEVDRIENENPRI-----EVRGIGQIMKEMSADQAT 318
 DB 766 KRYTEEREERYSYLRYTGTPIPDNDNKNNSQLVYVAGTVEGTDISLKFEEIDLTS 825
 QY 319 RIKYASKYAO-----SANWKNSIGMNRGLARDVIGRRABERAFADIRKN----- 366
 DB 826 RPAHGKTEQGLSPKSKPRATDSGAMSHKLEKADLL--KAIDQOLIAN--VHSNDVFEVY 882
 QY 367 --GKSAVYGDVLSSLEKAYKEGAKANREMTYLSSETLFGSTEYVRRQAFNALATNPDAHA 424
 DB 883 DRASDATITDRNGKYVFADKGS-VTLPTQPOVQEFLLSGHVRYKE-----KPIQNO 934
 QY 425 GLIKSLDDKY-----KDYLPFL-DRKYLPAMLDIVRRIRIPADKLPIEFKNVIDKK 473
 DB 935 A--KSVADVEYVQETPLNDDDFRPLGDKTKLKTLL--AIGDITTSQELLAQAOSILNKN 990
 QY 474 FKGDPFKYADVFYDSKVPYSDFKHAMLSMDEKFAKAIENDPAVELSKSVYAAARAQ 533
 DB 991 HPG---YTIYERDSIVTHDNDIFRTILPMDQEFYRKKNRQOAVRINKKSGLINEINN 1046
 QY 534 ADAMANA-YAIEKSK-----RLFAGILREMYPGARALPSDANFT---MRMSYGS 578
 DB 1047 TDLISKYVVLKKGEPYDPFRSHLKFTIKYVVDVDTMELKSEDLALASENLDLFRD 1106
 QY 579 KGYEPODGA--WYN-----YHTTGKGVLEKODPKSDEFAVOENILDLFR--TKNYG 625
 DB 1107 --YDPPDKAKLLYNNLDARGIMDYTLTGK-----VEDNHDDTNRIITYVMG 1150
 QY 626 R--YAENGQLHTAF 637
 DB 1151 KRPEGENASYHLAY 1164

RESULT 7
 US-08-488-940-17
 Sequence 17, Application US/08488940
 Patent No. 5854049

GENERAL INFORMATION:
 APPLICANT: Reed, Guy L.
 TITLE OF INVENTION: PLASMIN-RESISTANT STREPTOKINASE
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA

COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/488,940
 FILING DATE: 09-JUN-1995
 CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
 NAME: Fraser, Janis K.
 REGISTRATION NUMBER: 34,819
 REFERENCE/DOCKET NUMBER: 05433/009001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154

INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1194 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-488-940-17

Query Match 3.3%; Score 122.5; DB 2; Length 1194;
 Best Local Similarity 19.3%; Pred. No. 0.035;
 Matches 130; Conservative 99; Mismatches 254; Indels 191; Gaps 35;

QY 75 GITVSDQGL-----IFTHHCGYGAIGOSQTVVDHDLRDGFVSRMGELPIP---GLS 125
 DB 571 GDTITSQELLAQAOSILNKNHPGTYIERDSSI---VTHDNDIFRTI---LPMDOEFYR 624
 QY 126 VYLRKIVAVTDKVEGOLGIDEMERLKAQEVCOELAKKE----- 167
 DB 625 VKNRQAVYRINKK-----SGLNEEINNTDLISEKYYVLKKGEPYDPFRSHLKFTIY 679
 QY 168 -NADENQOLIVEPFYSNNY---FLIVYDFKDYRMVFPAPSSVGF-----GDTDN 216
 DB 680 VVDVNTLLKSQLTLASERNLDFRLYDPRDKAKLLY---NNLDAFGIMDYTLTGKVD 736
 QY 217 WMPVRHTGDSVFRVYAGADNRPAREYSKDNKPKPYFAAVSMQGYKADYAMTIGFPGS 276
 DB 737 ----NHDDTNRIITYVMG---KRP---EGENASYHLAY-----D 765
 QY 277 TDYR-----LTSWGEVDRIENENPRI-----EVRGIGQIMKEMSADQAT 318
 DB 766 KRYTEEREERYSYLRYTGTPIPDNDNKNNSQLVYVAGTVEGTDISLKFEEIDLTS 825
 QY 319 RIKYASKYAO-----SANWKNSIGMNRGLARDVIGRRABERAFADIRKN----- 366
 DB 826 RPAHGKTEQGLSPKSKPRATDSGAMSHKLEKADLL--KAIDQOLIAN--VHSNDVFEVY 882
 QY 367 --GKSAVYGDVLSSLEKAYKEGAKANREMTYLSSETLFGSTEYVRRQAFNALATNPDAHA 424
 DB 883 DRASDATITDRNGKYVFADKGS-VTLPTQPOVQEFLLSGHVRYKE-----KPIQNO 934


```

: LENGTH: 1577 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Streptococcus salivarius
US-08-793-824-2

```

```

Query Match 2.9%; Score 109; DB 2; Length 1577;
Best Local Similarity 20.7%; Pred. No. 1;
Matches 145; Conservative 95; Mismatches 262; Indels 198; Gaps 39;

```

```

QY 46 GTFLPLDLSL-YSPDKPSTIANAVVFSGGCGITVSDGLFTNHCYGA-IGQSQSYVDH 103
DB 690 GTFLTMELKQAEF--LYNA-----DIANDKKRYQYNIPAAVATMLTNKSDITR 737
QY 104 DYLRDGFV--SRMGELPIPLGLSVKYLKIVKVTDRKVEGOLKGTDEMERLRAQEVQC 161
DB 738 VYVGDLFTDDGQYMAEKSPYNAIDALLRAIK-----YVAGQDMKYTKLNGYEIMS 790
QY 162 ELAKKENADE-NOLCIYEPFYSNNEFLIYDVFKDYRVPAPPSVSKTEGDTDNMMP 220
DB 791 SVRTGKAEEMOLGTAE--TRNOGMLVLTANRPDKML-----GANDRLV-- 833
QY 221 RHTGDFSVFRYAGADNRPAEYSKDNKPYKPYFAA-----VSMQ 260
DB 834 -----VNMGAHKK-----NOAYRPLLSKSTGLATYTKDSVPAGLVRYTDNQ 876
QY 261 G--YKADYVA--MTIGFGSTDRYLTWSGVEDRIENENNRPIEVRIGIKQ--IMKRAMS 313
DB 877 GNLFETADLDAGHSTVEVS--YLAVWVPVYGASENQ-DARTKASSPKKGEQFESSAA 931
QY 314 ADQAT-----RIKYASKY-----AQSANYKNSIGMNRGLARLDVIGR-KRAEBR 357
DB 932 LDSQVYIEGFSNFQDFVKTPSQYTNRYIAQNAKLK-----EMGITSFEFAPQYVSSODG 986
QY 358 AFADWIRKNG-----KSAVVG--DVLSSLEKAKKEGAKANREMTYSETLFG 402
DB 987 TFLDSTIENGAFEDRDIAMSKNNKYGSLKDLALRALHAEISALAD--WPPDQIYN 1044
QY 403 --GTEVVRFAQFANALATNPDAHAGILKSL-DMYKDYPLSLDRKVLPAMLDIYRRRI 459
DB 1045 LPKREYVT-ASRTNSYGH-PRPNAEIYNSLYAAKTRTEGNDFOGKYGGAFLDELKAKYPA 1102
QY 460 DKLPDIKRNV-IDKFKGDTKRYADEVFEDKSVVPYSDKFAHMLKSMDEKFAKAIKEDPA 518
DB 1103 -----IFERVOISNGRKLT-----NEKITOMSAKY----- 1128
QY 519 VELSKSVI--AAARAIGADAMANAATIEKGRLEFFAGIREMYPG--RALPSDANFTMRM 573
DB 1129 --FNGSNIQTGARYVADNATNOYFSYKAGQTLPRKOMETITSGSFRVGD--V 1180
QY 574 SYGSIKGEY-----ODGA--WYNYHTTGKGVLEKODPKSDFAVQENIIDLFETKNYGR 626
DB 1181 QYLSIGGLAKNTFTIYGANQWYFEDKNGNMYTEGEQYIDGKKYFFLNDGLD---RHVLR 1237
QY 627 YAEKGQIHI-----AFLSNNDITGNSGSPVFDKNGRL 659
DB 1238 QGSDGHVYVYDDPKGVQAFNGFYDFAGPPODVRYFEDNGOM 1277

```

```

RESULT 10
US-08-335-844A-23
: Sequence 23, Application US/08335844A
: Patent No. 6066503
: GENERAL INFORMATION:
: APPLICANT: GRAHAM, MARGARET
: APPLICANT: SMITH, TREVOR STANLEY
: APPLICANT: MUNN, EDWARD ALBERT
: APPLICANT: KNOX, DAVID PATRICK
: APPLICANT: OLIVER, JOANNA JANE
: APPLICANT: NEWTON, SUSAN ELIZABETH

```

```

: TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
: TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
: TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS
: NUMBER OF SEQUENCES: 73
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Rothwell, Figg, Ernst & Kurz
: STREET: Suite 701-E, 555 Thirteenth St., N.W
: CITY: Washington
: STATE: D. C.
: COUNTRY: U.S.A.
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/335,844A
: FILING DATE: 09-JAN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB PCT/GB93/00943
: FILING DATE: 06-MAY-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9209936
: FILING DATE: 08-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: WALKER, Barbara W.
: REGISTRATION NUMBER: 35,400
: REFERENCE/DOCKET NUMBER: 1181-223A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)783-6040
: TELEFAX: (202)783-6031
: INFORMATION FOR SEQ ID NO: 23:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 972 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-335-844A-23

```

```

Query Match 2.9%; Score 106; DB 3; Length 972;
Best Local Similarity 19.9%; Pred. No. 0.87;
Matches 103; Conservative 71; Mismatches 223; Indels 120; Gaps 21;

```

```

QY 89 HCGYGAIGQSSTVDHYLDLDRGFSRTMGELPIPLGLSVKYLKIVKVTDRKVEGOLKGTID 148
DB 383 HWPFGNLTTLKMWDDTLMNLFGEFA--TEVEYLGMDIEISHNFR--TDFFLIDGMDRGMA 438
QY 149 EME-----RLKAEVQC-----ELAKKENADENOLCIPEFFYS 182
DB 439 DSAASSHPLSRFLDKAEVVAEAFDDISYANGASVYLMRALIGEDNRNAVVOYLKRFYS 498
QY 183 NNEFYFLIYVDYFDK-VAMVVAAPSSVGKFGGDTDNMMPRHGTGDFSVFRYAGADNRPAAE 241
DB 499 SNAQADLMVNFENVAVGVPDGNVAKIQDFTQWYMGCPYVYKKEEFNATLAKYTQS 558
QY 242 YSKDNKPKPYFAVAVSMOGYKADYMTIGFP-----GSTDRYLTSGVEDR---IEN 292
DB 559 RYKTNKD-----ALEPEKRYNPKYGFKMDVPLMYOEGNSKEVKRTWLKRDDEPLYLVNV 611
QY 293 ENNPRIEVRIGIKGIMKEAMSADQATRIKYASKAQAQSANVWKNSIGMNRGLARLDVIGRK 352
DB 612 NRDTSLVAVNADRHEGYQO-----NYDANGKK-----ITKOL 643
QY 353 RAEEARAFADWIRKNGSAVYGVYDLS-----LEKAYK--EGAKANREMTYSETLFGG 403
DB 644 KKHKVEGEPTR-----NAIISDAFAAATIDAIDVEYFELLELYAKNNEEFLPKWEALSGM 699
QY 404 TEVVRFAQFANALATNPDAHAGILKSLDDKYYKDYPLSLDRKVLPAMLDIYRRRIADKLP 463
DB 700 FAVLKF--FGNEPETKP-ARAYMMSILEPMYMK--SSIDYIVKAYLDLDTLFTKINTQK-- 752

```

QY 464 DIFKNVIDKFKGDTKKYADVFVDSKSVYSDKFHAMLKS-----D 505
DB 753 DIIDAYCSIGSKDCIKQYKIDIFYDE-VMPCKKAGEATKCVASPLRANVYCYGVQEG 811
QY 506 KEKFAKA-----LEKPAVELSKSVIAAARAIADADAMA 538
DB 812 EBAFEKVMGLYLAED--VQLEKGIILFKALACHKVDYTA 846

RESULT 11
US-08-811-583-2
Sequence 2, Application US/08811583
Patent No. 6218142
GENERAL INFORMATION:
APPLICANT: Massenegger, Michael
APPLICANT: Riedel, Leonhard
APPLICANT: Schiebel, Winfried
APPLICANT: Sanger, Heinz
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
TITLE OF INVENTION: RNA-DIRECTED RNA POLYMERASE (RDRP)
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/811,583
FILING DATE: 05-MAR-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: MPG-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1114 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-811-583-2

Query Match 2.8%; Score 103.5; DB 4; Length 1114;
Best Local Similarity 19.8%; Pred. No. 1.9; Mismatches 229; Indels 207; Gaps 30;
Matches 127; Conservative 78;

QY 92 YGAISQSTVDHDIYLDGVSRTMGE-----ELPIGILSVY-----LRKIYK 134
DB 556 YGKYGKGVGVDPDSSMKLSLRKSMKSYESDNKIKLDVLGMS-KYQPCYLNKRLITLSTIG 614
QY 135 YVDK-V-EGOLKITDMEERL-----RKAQEVCOELAKKENADENQICI-----VEPFYS 182
DB 615 VQDEVYLEQKOKKBAVDOLDAILHDSLKACQALMLMSGENTILKMLNCGYKPDAPFLS 674
QY 183 NNEFYLIYDVFK-----DVR---MYFAPSSV-----CK 209
DB 675 -----MMQTRASKLIDLRTSRIFIPNGRTMMGCLDESRTLEYGVQVFTAGHGE 728
QY 210 FCGDITDMMMPRHGTGDSVFYRYACADNRPAYSKDNKPYKPYFAAVSMOGYKADVDYAM 269
DB 729 FSDDLHPFNNSSTNSNFTLKGNVYVAKNPKLHPGDIRVLKAVNVRALH---HMPD----- 781

QY 270 TIGFP-----GST---DRYLTSWGYEDRIENENNRIEVRGIKOGIMKEAMS- 313
DB 782 CVYFPGQKRRPHRPNBCSGSLDGLDIFYVCW-----DQMIPRQVQPMPEYPPAPSI 832
QY 314 -ADQATRIYKASKYQASANYTKN-SIGMNRGLARLDVIGRKAERAFADWIRKNG----- 367
DB 833 QLDHVTJEEVEBEY--FTNYIVNDSLGI--IANAHVFADEPDMAMSDPKKLAELFS 887
QY 368 -----KSAVYGDVLSLE-KAYKEGAKANREMYLSETLFGCTEVRVPAQANALATNP 420
DB 888 IAVDFPKTVPAEIPISQLRPKETPPDMKPKDKTSIISERVIG----- 929
QY 421 DAHAGLNSLDCKYKDYLPSLDRKYLPAMLDIYRRIRPADKLPDIFKNVIDKFKGDTKK 480
DB 930 -----KLFRKVKDKAPQAS-SIATFTROYARSDADMEVDEFDYIDEAF--DYKT 978
QY 481 YADVFEDKSVYVYSDKFHAML-----KMDKEKFAKALEKPAVELSKSVIAAARA 531
DB 979 EYDNKLGMLMDYGIKTEKEILSGIMASKTFDRRKAELTS-----VAARA 1026
QY 532 IOADMANAYALEKGRLEFAGLRREYPCRALPSDANFTMRSGYSIKGYEPQDCAWYN- 590
DB 1027 LKKEAR-----AMFRRR---NDIDDMLPKASAWYHV 1054

QY 591 -YHTTGKGY---LEKDPKSDFAVQENIIDLFTKNYGR 626
DB 1055 TYHPTWYGCYNOGLKRAHPISEPMCVYDQLIQIKDKAKNR 1095

RESULT 12
US-09-514-302-3
Sequence 3, Application US/09514302
Patent No. 6338959
GENERAL INFORMATION:
APPLICANT: HATADA, Yoji
APPLICANT: IGARASHI, Kazuaki
APPLICANT: OZAKI, Katsuya
APPLICANT: ARA, Katsutoshi
APPLICANT: KAWAI, Shuji
APPLICANT: ITO, Susumu
TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PHOSPHATASE AND
TITLE OF INVENTION: ALKALINE ALPHA-AMYLASE ACTIVITIES
FILE REFERENCE: 2173-105P
CURRENT APPLICATION NUMBER: US/09/514,302
CURRENT FILING DATE: 2000-02-28
EARLIER APPLICATION NUMBER: 08/952,084
EARLIER FILING DATE: 1997-11-10
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO: 3
LENGTH: 833
TYPE: PRT
ORGANISM: Bacillus sp.
US-09-514-302-3

Query Match 2.8%; Score 102.5; DB 4; Length 833;
Best Local Similarity 19.0%; Pred. No. 1.4;
Matches 108; Conservative 71; Mismatches 199; Indels 191; Gaps 25;

QY 231 YVAGADNRA-----EYKDNKPYKPYFAAVSMOGYKAD-DYAMTITGPGSTDRRLTSMG 285
DB 298 YVAGVKNLPITAIIDEGFNHESGATLEVOARTITGEKADFDWDQSVVYFMLTDRFF----- 353
QY 286 VEDRIENENNPR-----IEVRGIKQ-----GIWKEAMSADQATRI 320
DB 354 --DGGSSNNDPFGIGTYDTSKSGTYGGDFKGTQGLDYLDELGITWITISPV-----VDNI 407
QY 321 KYASRYAQSAN-----YWRNSIG-----MNRGL-ARLDYI--- 349
DB 408 KFDVHSECPDPYVAYHGYWADNFGELNPHRGSMADFHEMIDAHERGIKIMVDVVLNH 467
QY 350 ---GKKRA-----EERA-FADWIRKNGKSAVYGDVLSLEKAYKBGAKANREMT 394

Db 468 TGCGKPCGSSSVANPEPTEDEDRARFDGMLRDGSGGEVRGE-LAGLPDELTEPNPDRVOV 526
QY 395 Y-----LSETLFGGT-----EVVFAQFANALATNPDAHAGILSLDDKK 435
Db 527 QWOTDMEKSRKTAKGNTIDYFRVDYKHYVEDTWMFAFKNALTAMPEHKLIGAMGANN 586
QY 436 DYLPISDRKVLPMALDIVRRIRPADKLPIEFKNVIDKKFKGDKTKYAAFEVDEKSVVPSD 495
Db 587 DDGGLNSGMDLDF-----DKNYARDAPOQLDAVOQ 622
QY 496 KPHA-----MLKSMDEKFAKAEKDPAVELSKSVIAAARLQADAMANAYA 542
Db 623 KLEARNSKLNTATLQOFLSHDEDRFEVEED---LGRYVAAASLQTLA----- 670
QY 543 IEKGRLEFAGLREMYPGRALPDSANFTMMSYGSIKGYEPQOGANYHTTGKGVLEQ 602
Db 671 --KGQVITYGEEGLGPK---ND-----YPTYTNQN-MPWD 702
QY 603 DPKSDEFAVOENILDFRTKNYGRYAENGOLHIAFLSNNDITGNSGSPYEDKNGRLIGL 662
Db 703 DVDGNELLEHYOKLAFRNPNPTFAKGRKVA---GSDSEGYLLFSRTYGENSVYVGL 759
QY 663 AFDGNEMASGDIIEFPDLORTISVDIRY 691
Db 760 ----NTEAAKDVTLNFGSSSEAVT-D-RY 783

RESULT 13
US-09-514-302-2
; Sequence 2, Application US/09514302
; Patent No. 6338959
; GENERAL INFORMATION:
; APPLICANT: HATADA, Yuji
; APPLICANT: IGARASHI, Kazuaki
; APPLICANT: OZAKI, Katsuya
; APPLICANT: ARA, Katsutoshi
; APPLICANT: KAWAI, Shuji
; APPLICANT: ITO, Susumu
; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PHOSPHATASE AND
; TITLE OF INVENTION: ALKALINE ALPHAMYLASE ACTIVITIES
; FILE REFERENCE: 2173-105P
; CURRENT APPLICATION NUMBER: US/09/514,302
; CURRENT FILING DATE: 2000-02-28
; EARLIER APPLICATION NUMBER: 08/952,084
; EARLIER FILING DATE: 1997-11-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1938
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-514-302-2

Query Match 2.8%; Score 102.5; DB 4; Length 1938;
Best Local Similarity 19.0%; Pred. No. 5.8;
Matches 108; Conservative 71; Mismatches 199; Indels 191; Gaps 25;
QY 231 VYAGADNRPA-----EYKSKDKPKRYFAAIVSQGYRAD-DYANTIGFPGSTDRYLTWSG 285
Db 298 VTGAGVKMLPTTALIDFEGNRHGSATLEVOARITTEKADPDQSVYVFWLDRFF----- 353
QY 286 VEDRIENENNR-----IEVGIRKO-----GIWKAMASDAQTRI 320
Db 354 --DQDSNNPDRHIGYITSSSGTYGGDFKGTORLDYDELGINTWISPV---VDNI 407
QY 321 KYASKYAQSAN-----YMKNSIG-----NMKGL-ARLDVI--- 349
Db 408 KFDYRHSRGEDTPTYYAHGWMADNFGELNPHFGSMADFHMDIAHHRGKIKINVDVNLH 467
QY 350 ---GRKA-----EERA-FADWIRKNGKSAVGYDLSLEKAYVEGAKANREMT 394
Db 468 TGCGKPCGSSSVANPEPTEDEDRARFDGMLRDGSGGEVRGE-LAGLPDELTEPNPDRVOV 526

QY 395 Y-----LSETLFGGT-----EVVFAQFANALATNPDAHAGILSLDDKK 435
Db 527 QWOTDMEKSRKTAKGNTIDYFRVDYKHYVEDTWMFAFKNALTAMPEHKLIGAMGANN 586
QY 436 DYLPISDRKVLPMALDIVRRIRPADKLPIEFKNVIDKKFKGDKTKYAAFEVDEKSVVPSD 495
Db 587 DDGGLNSGMDLDF-----DKNYARDAPOQLDAVOQ 622
QY 496 KPHA-----MLKSMDEKFAKAEKDPAVELSKSVIAAARLQADAMANAYA 542
Db 623 KLEARNSKLNTATLQOFLSHDEDRFEVEED---LGRYVAAASLQTLA----- 670
QY 543 IEKGRLEFAGLREMYPGRALPDSANFTMMSYGSIKGYEPQOGANYHTTGKGVLEQ 602
Db 671 --KGQVITYGEEGLGPK---ND-----YPTYTNQN-MPWD 702
QY 603 DPKSDEFAVOENILDFRTKNYGRYAENGOLHIAFLSNNDITGNSGSPYEDKNGRLIGL 662
Db 703 DVDGNELLEHYOKLAFRNPNPTFAKGRKVA---GSDSEGYLLFSRTYGENSVYVGL 759
QY 663 AFDGNEMASGDIIEFPDLORTISVDIRY 691
Db 760 ----NTEAAKDVTLNFGSSSEAVT-D-RY 783

RESULT 14
US-07-854-596B-40
; Sequence 40, Application US/07854596B
; Patent No. 5434073
; GENERAL INFORMATION:
; APPLICANT: Dawson, Keith M
; APPLICANT: Hunter, Michael G
; APPLICANT: Czaplinski, Lloyd G
; TITLE OF INVENTION: Proteins and nucleic acids
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. John J. McDonnell
; STREET: Ten South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/854,596B
; FILING DATE: 03-JUN-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDonnell, John J
; REGISTRATION NUMBER: 26,949
; REFERENCE/DOCKET NUMBER: 92,337
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-854-596B-40

Query Match 2.7%; Score 101.5; DB 1; Length 747;
Best Local Similarity 19.3%; Pred. No. 1.5; 254; Indels 183; Gaps 33;
Matches 127; Conservative 94; Mismatches 183; Gaps 33;
QY 75 GITVSDGL-----YFTNHGCGYGAIOSQTVHDYLRDGFYSRTWGEPLIP---GLS 125

```

Db 177 GDTITSOELLAQOSILNKHGTYTYERSSII--VTHNDIFRTI---LPMQDEFTYH 230
OY 126 VYLLKRYKYVDKVGQKLGIDEMERLKAQVCOELAKKE----- 167
Db 231 VANRQDAVEINKK-----SGLNEEINNDLISEKYVLLKGEKYPDPDRSHLKLFTTKY 285
OY 168 -NADENOLCIYEPFYSNNY---FLIYVDVFNKDVMPVAPSSVGKFG-----GDTDN 216
Db 286 VQVNTNELLKSQLLTJASERNLDPRDLDVDPKAKLTY---NNLDAFGIMDYTLGKVED 342
OY 217 WMMPHRTGDSFERYACADNPAEYSKDNKPKYVYPAVSMQYKADDTAMTIGFPGS 276
Db 343 ---NHDDTNRIITYMG--KKRPEGEMAS-----YHLAVELQGVPRGSQLVYVAGT 389
OY 277 TDRYLTSMGVEDRIENENNPRIEVRGIKQIWKEMSADQATRIYASKYAO-----SAN 331
Db 390 -----VEGTNDISLKEFEIDLTRPHAGKGTGGLSPKSKP 426
OY 332 YKNSIGMNRGLARLDVIGRKRAERAPADWIRKN-----GKSAVYGDVLSLEKA 382
Db 427 FATDSGAMPKLEKADLL--KAIQEQLIAN-VHSNDYFEVIDFASDATTDRNGKVVFA 483
OY 383 YEGKAKANREMYLSETLFGTEVYVRFQAFANALATNPDAHGI---LKSIDDKY----- 434
Db 484 DKDG-----SVTL--PTQVQOEFLLSGHVVRPRKPKPIQNAKSQVVEYTVQFT 531
OY 435 -----KDYLPSTL-DRKVLPRAMLDIVRRIRPADKLPIEFKNVIDKFKGDPKKYADFVEDK 488
Db 532 PLNPDDDFRPGJADKDKLKLTL--AIGDITTSOELLAQOSILNKHG-----YTIYERNS 585
OY 489 SVYVPYSDKFHAMLKSMDKEKFAKAIENDPAVELSKSVIAAARAIQADAMANA-VYIEKGK 547
Db 586 STVTHNDIFRTITLPMQDEFTYHVKNRQDAVEINKKSGLINEEINNTDLISEKYVLLKGE 645
OY 548 -----RLFPAGLREMYGRALPSDANFT---MRMSYSGIKGYEPDQGA--WTN- 590
Db 646 KRYDPDSHKLFTITIKYVDVNTNELKSEQLLTJASERNLDPRDL--YDPRKAKALLINN 703
OY 591 -----YHTTGKGVLEKQDPKSDFAVOENILDLFR--TKNYGR--YAENGOLHIA 636
Db 704 LDAFGIMDYTLGK-----VEDNHDDTNRIITYMGKRPKEGNASVHLA 747

RESULT 15
US-08-417-492-2
; Sequence 2, Application US/08417492
; Patent No. 5750872
; GENERAL INFORMATION:
; APPLICANT: Bennett, Alan B
; APPLICANT: Brummell, David A
; APPLICANT: Grantz, Alexander A
; TITLE OF INVENTION: Nucleic Acids Encoding Ascorbate Free
; TITLE OF INVENTION: Radical Reductase and Their Uses
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Stewart Street Tower
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,492
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L
; REGISTRATION NUMBER: 34,774

```

```

; REFERENCE/DOCKET NUMBER: 2307E-586US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-417-492-2

Query Match 2.7%, Score 101; DB 1; Length 433;
Best Local Similarity 21.3%, Pred. No. 0.66;
Matches 102; Conservative 62; Mismatches 168; Indels 146; Gaps 23;

OY 220 RYVACADNRPAEYSKDNKPKYVYPAVSMQY-----KADYAMTIGPGRDRLT 282
Db 22 REFAGQVKGPGELATISKAEVAPYRPPALSKAYLFPEGAARLPGFHVCGSGG--ERQLP 79
OY 283 SMGVEDRIENENNPRIEVRGIKQIWKEMSADQATRIKYASKYQASANYKNSIGMNRG 342
Db 80 EMYAE-----KGISLITSTELVAKADLAK--TVSAAQESFRKYQTVIATGTT 125
OY 343 LARLDVIGRKRAERAF-----ADWI-----RKNGKSAVYGDVLSLE----- 380
Db 126 VKLSDDFGVQAGDSKNIFYLREIDADQDLVEALKAKKNGKAVVGGYIGLELSAVLRIN 185
OY 381 -----KAKKEGAKANREMYLSETLFGTEVYVRFQAFANALATNPDAHA- 424
Db 186 NIEVNMVYPEPWCMPLRTETGIAAFYEGYKKN--GVNIKK-----GYAVAGDTHPN 236
OY 425 GILKSIDDKYKDYLPSTLDRKVLPRAMLDIVRRIRPADKLPIEFKNVIDKFKGDPKKYADF 484
Db 237 GEVKEVKKL-----DGRVLEA--DLVVGVGARPLTLEKGVVEE--KGGITDA-- 283
OY 485 VEDKSVVP-----YSDKFHAMLKSMDKEKFAKAIENDPAVELSKSVIAAARAIQADAMANA 540
Db 284 -FFKTSVPDYAVAGVATFPPLKMYNEIRREHVDHS-----RKSADQAVKAI----- 329
OY 541 VAIERKRLFFAGLREMYGRALPSDANFTMRMSYSGIKGYEPQGANVNY-HTTGKGVL 599
Db 330 FASEQKGSVDEYDILPYTSRAF-----DLSQGYDANNGEYVL 368
OY 600 -EKQDKSDEFAVOENILDLFRTKNYGRY-AENGOLHIAFLSNNDITGNGSSGPVFDK 635
Db 369 FGDADPNs-----ATHKFGQYWKDGKIVGAFI-----ESGSPENK 405

```

Search completed: October 17, 2002, 22:41:07
Job time : 32 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:40:26 ; Search time 2436 Seconds
(without alignments)
6116.459 Million cell updates/sec

Title: US-10-008-355-2
Perfect score: 3719
Sequence: 1 MOWKLSILGALLIGASG.....LFMDKMGCCPRLLIQLKLI 712

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n model -DEV=tlh
-Q/cgn2_1/USPTO_SPOOL/US10008355/runat_16102002_121153_27693/app_query.fasta_1.903
-DB=GenEmbl -QWMT=fastap -SUFFIX=rg -MINMATCH=0.1 -DOOPCL=0 -DOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCAALIGN=200 -THR_SCORE=DCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355 @CGN1.1.1601 @runat_16102002_121153_27693 -NCPu=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEORDER -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_lm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
C 1	907.5	24.4	10689	1 AE004008	AE004008 Xylella f
C 2	186.5	5.0	253217	2 AC016590	AC016590 Homo sapi
C 3	146.5	3.9	10889	1 AE006549	AE006549 Streptoco
C 4	144.5	3.9	15569	1 AE000733	AE000733 Aquifex a
C 5	137.5	3.7	162595	2 OSJN00169	AL662969 Oryza sat
C 6	133.5	3.6	4004	1 MSP207AAM	X55799 M.sp 207 al
C 7	133	3.6	10292	1 AE008428	AE008428 Streptoco
C 8	133	3.6	11362	2 SPNEU1923	AL449945 Streptoco
C 9	131.5	3.5	14169	1 AE009858	AE009858 Pyrobacul
C 10	130	3.5	2652	6 AB1135	AB1135 Sequence 1
C 11	130	3.5	2652	6 AR065782	AR065782 Sequence
C 12	130	3.5	2652	6 E26074	E26074 Novel vials.
C 13	129	3.5	4141	1 CLOCOLA	D13791 C. perfring
C 14	129	3.5	6004	6 BD003707	BD003707 Polynucle
C 15	129	3.5	11280	1 AE007366	AE007366 Streptoco
C 16	129	3.5	26703	3 CBRG45011	AC084652 Caenorhab
C 17	128.5	3.5	291150	1 AP003135	AP003135 Staphyloc
C 18	128.5	3.5	342600	1 AC003363	AC003363 Staphyloc
C 19	127	3.4	193188	2 AC006884	AC006884 Caenorhab
C 20	126	3.4	3084	1 LAU62096	U62096 Lactobacill
C 21	124.5	3.3	9389	1 AE008276	AE008276 Agrobacte
C 22	124.5	3.3	11199	1 AE009339	AE009339 Agrobacte
C 23	124	3.3	36380	3 AF106581	AF106581 Caenorhab
C 24	123.5	3.3	2142	1 AF039313	AF039313 Moraxella
C 25	123.5	3.3	10296	1 U67538	U67538 Methanococ
C 26	123	3.3	10181	1 AE001761	AE001761 Thermotog
C 27	123	3.3	340806	1 NMA122491	AL162751 Neisseria
C 28	122.5	3.3	7009	14 AF296095	AF296095 Porcine t
C 29	122.5	3.3	7858	1 U35629	U35629 Lactococcus
C 30	122.5	3.3	10826	1 SPH49397	U49397 Streptococc
C 31	122.5	3.3	239050	1 AL596169	AL596169 Listeria
C 32	122.5	3.3	299850	1 AP001514	AP001514 Bacillus
C 33	122	3.3	10698	1 AF218939	AF218939 Bacillus
C 34	122	3.3	301450	1 AP003185	AP003185 Clostridi
C 35	122	3.3	333050	1 AL596168	AL596168 Listeria
C 36	121	3.3	3225	3 AF438184	AF438184 Toxoplas
C 37	121	3.3	3516	3 AF438183	AF438183 Toxoplas
C 38	121	3.3	4419	3 AF006628	AF006628 Toxoplas
C 39	121	3.3	4535	3 AF006627	AF006627 Toxoplas
C 40	121	3.3	11073	1 AE006589	AE006589 Streptoco
C 41	120	3.2	10380	1 AE000932	AE000932 Methanoba
C 42	118.5	3.2	291050	1 AP000982	AP000982 Sulfolobu
C 43	118.5	3.2	339485	8 AF172282	AF172282 Oryza sat
C 44	118	3.2	11069	1 AE004124	AE004124 Vibrio ch
C 45	118	3.2	39937	7 VTG6	V01146 Genome of b

ALIGNMENTS

RESULT 1
LOCUS AE004008 10689 bp DNA linear BCT 15-JUN-2001
DEFINITION Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION AE004008 AE003849
VERSION AE004008.1 GI:9106961
KEYWORDS
SOURCE Xylella fastidiosa 9a5c.
ORGANISM Xylella fastidiosa 9a5c

Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.

1 (bases 1 to 10689)

REFERENCE
AUTHORS
Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bata,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R., Camargo,A.A., Camargo,L.E., Carraro,D.M., Carrier,H., Coutinho,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Fromme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Ho,P.L., Hohnselt,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P. and Marino,C.L.

TITLE
The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Nature 406 (6792), 151-157 (2000)

20365717
10910347

2 (bases 1 to 10689)

REFERENCE
AUTHORS
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bata,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Fromme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hohnselt,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.Y., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmieri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pequerio,J.B., Quaglio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tshabko,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Zedler,J. and Setubal,J.C.

TITLE
JOURNAL
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

FEATURES
source
1. 10689
Location/Qualifiers
/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"
/clone="9a5c"
complement(119. 367)
/gene="XF1871"
complement(119. 367)
/gene="XF1871"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84677.1"
/db_xref="GI:9106962"
/translation="MTWFLKEDKTEIIDEOKRIQIAKLISEAKIQAETHEVHTQKL MKETLKTTESKRWYPPVVGSSIIAAGATAAATLFIKLFH"
complement(430. 618)
/gene="XF1872"
complement(430. 618)
/gene="XF1872"
CDS

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84678.1"
/db_xref="GI:9106963"
/translation="MMEARATAPQGGIASDPQRTIDALQTVSHKRRRAANETASIIQLRQ IVLRLVLRQROKRRASP"
941. 1288
/gene="XF1873"
941. 1288
/gene="XF1873"
/note="similar to GI1479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF84679.1"
/db_xref="GI:9106964"
/translation="MSKKITPNIQTGMAAGFQFINKLVTPEPTIEKQRLMLSL CTLAERWQKMEEARATAPQGGIASDPQRTIDALQTVSHKRRRAANETASIIQLRQ VLIROOKRRASP"
1351. 1662
/gene="XF1874"
1351. 1662
/gene="XF1874"
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84680.1"
/db_xref="GI:9106965"
/translation="MTWFLKEDKTEIIDEOKRIQIAKLISEAKIQAETHEVHTQEM NAHINENMARQEMNARQKLIETQKSKESQWYPPVVASGLAAGATAAATLFIKLF H"
complement(1839. 2918)
/gene="XF1875"
complement(1839. 2918)
/gene="XF1875"
/note="similar to SP1P03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84681.1"
/db_xref="GI:9106966"
/translation="MLNLVTCAGPGNGKTLVAVDMILRIQIEIDKSLVKSQAVPRSYTD IEGDVEAVRRLTGYVDSAPEDKRTTQGSYIYDEAHRRFPACGRPRSRDDPRCDL DTHRRGGIDLMFVTQWPTKIHHELKRLVGEVHLNLRAGLQTAGIYKSRADDDYDI HOREKAEVEVKFPDRALYASSTLHVSFKFRIPKRWASLSCVYSGVIGLVFMI YSPDAHSEASVSAGASQASLRAAPASLSSSSIVSGMTYAVLLETESAPLTSGCV SSDRCRCFNTDGYOIDMSVYECRLASPLPFNVYHAYVYSSSAAVSSPSSPS SSVSYLSSVSASPLIGSSVPSH"
complement(2918. 3202)
/gene="XF1876"
complement(2918. 3202)
/gene="XF1876"
/note="similar to SP1P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 86.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84682.1"
/db_xref="GI:9106967"
/translation="MFSWLAMLLRNIIAGQIVYARGLVAGLALVTTVPLIPLVTSANL


```

QY 424 -----AlaGlyIleLeuLysSerLeuasp 431
||| |||||
Db 9230 GAAGGCGATATCAGAGCGGATCTGCCACCATGTGAAGCACTTAAGCAGATGAG 9171
QY 432 AsplysTyrLysAspTyrLeuProSerLeuaspArgLysVal ----- 445
||| |||||
Db 9170 -----CGACGTTACTTACCGAATAATGATCGCCAGATGACAGCACTACTGCGTCACC 9120
QY 446 -----LeuProAlaMetLeuaspLileValArgArgIleProAla ----- 459
||| |||||
Db 9119 GAATACACACAGCTTCCCT -----GTCAACACGCGTGTGCGTCGCAATTGAT 9075
QY 460 -----AspLysLeuProAspLilePheLysAsnValIleAspLysLysPheLys 475
||| |||||
Db 9074 GTTGTGTGGAGATGAGCATCCCAACACTCTCAAGCGCTTGGGACACCAAAATTAGC 9015
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerasp 495
||| |||||
Db 9014 AGTCTGAGGAAGG ----- 9000
QY 496 LysPheHisAlaMetLeuLysSerMetaspLysGluLysPheAlaLysAlaIleGluLys 515
||| |||||
Db 8999 -----CTCAATGTTTATATCCCGACCGCTCGACCGCTTGAAGACAGTCAA 8955
QY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGluAlaasp 535
||| |||||
Db 8954 GATCCAGCATCCGTTATGCTGTGCCATCATGCCAGCATTAAGATTAAGATTAAGCAG 8895
QY 536 AlaMetAlaAsnAlaTyrAlaIleGluLysGlyAspGluPhePheAlaGlyLeuArg 555
||| |||||
Db 8894 AATAAATCCGTACTGAGAACTACTTAACGGCGTCGATCTACTGCAAGCCTTACT 8835
QY 556 GluMetTyrPro -----GlyArgAlaLeuProSerAspAlaAsnPheThrMetArgmet 573
||| |||||
Db 8834 GACTACACACAGAGCATAGCGAAATTTGTTACCCAGACGCCACTCTCATTTACGATTT 8775
QY 574 SerTyrGlySerIleLysGlyTyrGluProGluAspGlyAlaTyrTyrAsnTyrHisThr 593
||| |||||
Db 8774 ACCTTCGCGCCAGCGTGAAGGATATTCACCAAGATGCGTGCAGTACACACCGCTTTCG 8715
QY 594 ThrGlyLysGlyValIleGluLysGluAspProLysSerAspGluPheAlaValGluGlu 613
||| |||||
Db 8714 ACACGCGAGGCTGTCATGCGCAAG ---AATACAGCGCTGCGACCATTCATTCACCGCAG 8658
QY 614 AsnIleLeuAspLeuPheArgTyrThrLysAsnTyrGlyArgTyrAlaGluAsn -----Gly 631
||| |||||
Db 8657 TCACGATTAATGCGATTAAGCAAGACAGATCAACCATTTAGCCGATCAGCATTTGCG 8598
QY 632 GluLeuHisIleAlaPheLeuSerAsnAspIleThrGlyLysAsnSerGlySerPro 651
||| |||||
Db 8597 ACTGTCCTGTAATTTCTCTGATCTTACATTTCTGCGCGCACTCGGATCGCA 8538
QY 652 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTyrGluAlaMet 671
||| |||||
Db 8537 GTCATGAGTGGCCACGCAAAATTTGGTGGTTAGCATTTGACGATTTGGAGAGCTCTC 8478
QY 672 SerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIleArgTyr 691
||| |||||
Db 8477 AGTAGTAATATGGGTGTTGATCTGTGATGACGCAACATTCAGTGAATGCGTTAT 8418
QY 692 ValLeuPheMetIleAspLysTyrGlyLysProAlaGluIleGluLeuLysLeu 711
||| |||||
Db 8417 GTGAGTGGATCATGACTGAAAGTTGACACACACACATCTACTAAAGAAATGAAATTG 8358

```

RESULT 2
 AC016590/c 253217 bp DNA linear HTG 20-APR-2001
 LOCUS Homo sapiens chromosome 19 clone CTD-3220F14, WORKING DRAFT
 DEFINITION
 AC016590
 AC016590
 AC016590.6 GI:13699590
 VERSION
 HTGS: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 KEYWORDS
 human.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Apr 20, 2001 this sequence version replaced gi:7711568.
 -----Genome Center
 Center: Joint Genome Institute
 Center code: JGI
 Web site: <http://www.jgi.doe.gov>

Project Information
 Center Project Name: 940643, BC905667
 Center clone name: CIRB-EL_3220F14

Summary Statistics
 Consensus quality: 207648 bases at least Q40
 Consensus quality: 225569 bases at least Q30
 Consensus quality: 232428 bases at least Q20
 Estimated insert size: 250510; agarose-ef estimation
 Estimated insert size: 249417; sum-of-contigs estimation
 Quality coverage: 9.03 in Q20 bases; agarose-ef estimation
 Quality coverage: 9.07 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 39 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1019: contig of 1019 bp in length
 * 1020 1119: gap of unknown length
 * 1120 2152: contig of 1033 bp in length
 * 2153 2252: gap of unknown length
 * 2253 3701: contig of 1449 bp in length
 * 3702 3801: gap of unknown length
 * 3802 5053: contig of 1252 bp in length
 * 5054 5153: gap of unknown length
 * 5154 6184: contig of 1031 bp in length
 * 6185 6284: gap of unknown length
 * 6285 7809: contig of 1525 bp in length
 * 7810 7909: gap of unknown length
 * 7910 8986: contig of 1077 bp in length
 * 8987 9086: gap of unknown length
 * 9087 10090: contig of 1004 bp in length
 * 10091 10190: gap of unknown length
 * 10191 11225: contig of 1035 bp in length
 * 11226 11325: gap of unknown length
 * 11326 12365: contig of 1040 bp in length
 * 12366 12465: gap of unknown length
 * 12466 14535: contig of 2070 bp in length
 * 14536 14536: gap of unknown length
 * 14537 14636: gap of 1854 bp in length
 * 14637 16488: contig of 1854 bp in length
 * 16489 16589: gap of unknown length
 * 16590 19109: contig of 2520 bp in length
 * 19110 19209: gap of unknown length
 * 19210 21841: contig of 2632 bp in length
 * 21842 21942: gap of unknown length
 * 21943 23735: contig of 1794 bp in length
 * 23736 23835: gap of unknown length
 * 23836 25861: contig of 2026 bp in length
 * 25862 25961: gap of unknown length
 * 25962 29117: contig of 3156 bp in length
 * 29118 29217: gap of unknown length
 * 29219 31907: contig of 2690 bp in length


```
gene
/organism="Streptococcus pyogenes M1 GAS"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
CDS
98..739
/gene="SPY1039"
98..739
/gene="SPY1039"
/feature="Best Blastp hit = no blast hits"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK33931.1"
/db_xref="GI:13622188"
/translation="MTKQIEKLAIDLPPFAATLYLDMIPERQEGYASYVYOTPKODI
OKRRIKQLRNFRLDPTSFYVMPYTRKTAEQDLKGLAYLLSTVANQOFAAY
NEDOMFOYLNRNGARAFVILQEGEFLMTGLILETYENEPRAIVAVQYLYLAD
MOKRLPSLSIMVQSESSKLFYQEGFVLAKGCKPQENSDVCQLVKYSKP"
802..1932
/gene="hemN"
/feature="hemN"
/feature="SPY1040"
802..1932
/gene="hemN"
/BC_number="1.3.3.3"
/feature="Best Blastp hit = pir||B69640 coproporphyrinogen
oxidase (EC 1.3.3.3) III, oxygen-independent hemN -
Bacillus subtilis >g11303805|dbj|BA12461.1| (DB4432)
Yger [Bacillus subtilis] >g11263496|emb|CAB14492.1|
(299117) coproporphyrinogen III oxidase [Bacillus
subtilis]"
/codon_start=1
/transl_table=1
/product="putative coproporphyrinogen III oxidase"
/protein_id="AAK33932.1"
/db_xref="GI:13622189"
/translation="MSKKPTSAVNHIPCTQICVYCDSPKVFQNPVAVYALKALQE
FDSYGRDLAKTLVIGGTPATITAKOLEVLNLENLDDLEETITANFEDLPE
KIAVLQSAVNRLISGVTFNNQOLQIGRSHNEQIYSTIANLKTAGFHNISIDLIY
ALPGQLDQKENVAKALADLIPHLISYLLLEHHTVENMKMRGLNPTEDLEEM
FEYIISEMANGFEHYEISNFTKPGFESRHNLMYQDNVEYFGCAGASGLNIGRON
RVPDQHYLAKEAGNARLNDEVLKREEMEEELFLGRKTKGVSIGRQEGKMSFEE
KGNIVRELQNGLLVAKDAFVMTKKGLFGDSVAERFLD"
2126..2695
/gene="SPY1042"
2126..2695
/gene="SPY1042"
/feature="Best Blastp hit = gb|AA071730.1| (U65643) acyl-ACP
thioesterase [Myristica fragrans]"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK33933.1"
/db_xref="GI:13622190"
/translation="MRSFLACIHKFSPSLKQKPEPIINFVIANFIMIRGGLV
DILAYFALLNPOTRKVATIPEDIVAFEPDFYKRLHVRKMPLEBDSIDRDYVRFD
IDMGHVNNSKTLDMWIDVGCFLKTHQPLKMTLKVKEVSPGQITSYHLDQLS
YQIITSQGQLNQAMLEIRAIKQTESETD"
2695..3459
/gene="SPY1043"
2695..3459
/gene="SPY1043"
/feature="Best Blastp hit = pir||H70023 N-acetyl-glucosamine
catabolism homolog ynfC - Bacillus subtilis
>g112635726|emb|CAB15219.1| (Z99120) similar to
N-acetyl-glucosamine catabolism [Bacillus subtilis]"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAK33934.1"
/db_xref="GI:13622191"
/translation="MPYKGYLDLQSTIYOGKNRIPAGRFKRLQDERGIPLYLTNN
TRTPBMVOSMLANQPHVETSIETITANMATVADVINDNRRGTAVVIGETGKSAIA
AAGYVEELENPAYVVGIDSOVYTEMALATLALQALIGTNPDLNIPTEGRIMPG
```

```
gene
AGALLNLEAATRVKPFVFGKPNALIMNKSLEVLQIRSEAVVWGDNYLTIDMAGION
DIATILVTGFTFRPEEVPLPIQPDHVLSSLEEMR"
3459..4091
/gene="SPY1044"
3459..4091
/gene="SPY1044"
/feature="Best Blastp hit = sp|P54589|YHCE_BACSU HYPOTHETICAL
29.5 KD PROTEIN IN GLPD-CSPB INTERGENIC REGION
>g117475050|pir||A69832 hypothetical protein ynfC -
Bacillus subtilis >g11239980|emb|CAB5688.1| (X69693)
hypothetical protein [Bacillus subtilis]
>g112633228|emb|CAB12733.1| (Z99108) ynfC [Bacillus
subtilis]"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK33935.1"
/db_xref="GI:13622192"
/translation="MVNTKLLCSWMLALAILITVSTWLMYPLEVDHLKEQVVF
MSKDAILHNYNGLNLITNPVYRLFEAFHSSADGLKHFADVKMLFULTQVFLGLL
YPLKFTQRLTKRPMLOKPLILALFALPEMIGMASPIGFHFPTLHOVLFGDS
SWLEPDKDSVIMILPEVFLHCFLEFMIVYEIILMSLVGLARMQRLANE"
4571..8677
/gene="SPY1046"
4571..8677
/gene="SPY1046"
/feature="Best Blastp hit = pir||H81982 hypothetical protein
NMA0631 [imported] - Neisseria meningitidis (group A
strain 22491) >g117379363|emb|CAB83920.1| (A162153)
hypothetical protein NMA0631 [Neisseria meningitidis]"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAK33936.1"
/db_xref="GI:13622193"
/translation="MDKRTSISLDITGNSVGAIVTDEKYSKFKVYLGNDNRHSIK
KULIGALDSDGSEATRLKRTARRRTYRRNRICYLOETISNEAKVDSDSFHLE
ESFLVEEDKRHRHPFGNIYDEVAHEHYPTIYHARKLIVSTQKADRLIYLLAH
MKFRFHPHIEDLNDSDVDKFLFQIQVQYNOUEENPISAGVDAAILIYARLSK
SRLEENLILQGEKKNGLFGLMIALSLGLTFNFSNPLADAKIQLSDKQYDDLD
NLALQDGOADYALFLAAKMLSDAILSLDILRVNTEITKPLASAMIKRYDEHODTL
IKAIYNOQLPERKYEFPDQSKNGACYIDGASQDEEYFKTPILEKDCREELLYK
LNRBDLKRQRTFDNGSTPHQIHGELMALIKRQEPFPLKDNKEKILITRIPY
YVGPLARGNSRFAMTRKSEETITPWNVEEVVDGSAQSFIERMTNEDKILNPKSVL
PKHSLYEYFYVNEILTKVYVTEGRKPAFLGDKKAIYDLKRTNKRKYQKLE
DYFKRIEEDFSVEISGVEDRFNASLGTIDYDLIKIKDKQFLENEEDLIEDIYVLT
LFEDRMTIEERLKTAVLFDQVMAOKRRRTYGRSGKRLINGIRKQSGKTIIDF
IKSDGFANRNPOLIHDSLTFREKIDOKRQVSGCDSJAEHIANIAGSPAIYKGIQT
VYVDELYVAKGRHKRPENIYEMARENOTQOGKNSRBRMRIRIEGRIELDSQILKE
HPVENTQLONEKLYLYLQNGRDYVVDQDLINRLSDVDVHIVQSFELKDSIDNKV
LRRSDKNRGSNDVSEEVYKRMKNYRQLNNAKLITOKFPNLKRAERGQSEIDKA
GFIKROLVETROITKHAQIILDSRMTKXDENDKILREKVYTLKSLVSDRKEQF
YKVEILNNYHHAHDVLYNAVGTALIKKPKLESEEVVDYVYVVRKMIASBOEIG
KATAKYFYSNINMNFPTTITLANGEIRKRLPIETNNGEGLYVMPKGRFAYRVYLS
MPQAVNPKRTVEQTOGFSKESITLPRNSKDLIARKDOPKRTVGGDSFTVAYSLV
AVEVNGSKRKLISVKEILGITIMERSSEFPKNIIDLEAAGYVREKDLILPKYSLV
ELENGKRMALASGELQKGNELALPSKYVNFYKLESGEDNEOKOLFQEOH
KHVDEIIEOISSEFSKRVILADANIKVYAVNKRDRKIRFOAENIHLFLTNLGA
PAAFKYEFTTIDRKRRTSKREVLDNMTLIHQSTIGLYEIRIDLSQLGQD"
8677..9546
/gene="SPY1047"
8677..9546
/gene="SPY1047"
/feature="Best Blastp hit = pir||B81299 hypothetical protein
Cj1522c [imported] - Campylobacter jejuni (strain NCTC
11168) >g116968946|emb|CAB73942.1| (A1139078) hypothetical
protein Cj1522c [Campylobacter jejuni]"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK33937.1"
/db_xref="GI:13622194"
/translation="MAGWRTVVVNRHSLKSYKNNHLIFKDAYKTELHLSEIDILLLE
```

```
gene
9543. .9884
/gene="SPY1048"
CDS
9543. .9884
/gene="SPY1048"
/Note="Best Blastp hit = emb|CA83918.1| (A1162753)
hypothetical protein NMA0629 [Neisseria meningitidis]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK33938.1"
/db_xref="GI:13622195"
/translation="MSRYRMRIIMFDMPTDAEERKAYRKRFKLLSGFTMHQFSI
YKLLNNTANNAMIGRLREHNPNKGNITLTVTEKQFARMYILHGERNNCLANSDE
LYFLGAPDES"
/gene="SPY1049"
9874. .10536
/gene="SPY1049"
9874. .10536
/gene="SPY1049"
/Note="Best Blastp hit = p1r|IT03910
```

Alignment Scores:

Pred. No.:	1.42	Length:	10889
Score:	146.50	Matches:	172
Percent Similarity:	31.98%	Conservative:	103
Best Local Similarity:	20.00%	Mismatches:	223
Query Match:	3.94%	Indels:	363
DB:	1	Gaps:	44

US-10-008-355-2 (1-712) x AE006549 (1-10889)

```
QY 35 AsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThrLeuProLeuAspSerLeu 54
|||:::||||| ||| ::::: ||||| |||
Db 6395 AATGAGAAATAATGAGATATCTTAGAGATATGTTTAAACATTGACCTTATTGAA--- 6451
QY 55 TyrSerPheAspIysProSerIleIleAsnIleValIlePheGlyGlyGlyCysThr 74
|||::: ||| ::::: |||
Db 6452 -----GATAGGAGATGATTGAGAAACCTTAAACATATGCTCAC----- 6493
QY 75 GlyIleThrValSerAspGlnGlyLeu-----IlePheThrAsnIleHisCysGlyTyr 92
|||::: ||| ::::: |||
Db 6494 -----CTCTTGATGATTAAGTGATGATAACACGCTTAACGCTTGCTTACTGGTTGG 6547
QY 93 GlyAlaIle-----GlnSerGlnSerThrVal 101
|||::: ||| ::::: |||
Db 6548 GGACGTTTGTCTCGAATAATGATTAAATGATTAGGATTAAGCAATCTGGCAAAACATA 6607
QY 102 AspHisAspIysPheLeuArg---AspGlyPheValSerArg----- 113
|||::: ||| ::::: |||
Db 6608 ---TTAGATTTTGAATGACATGTTTGGCAATGCAATTTTATGACGTGATCCAT 6664
QY 114 -----ThreGlyGluGluLeuProIleProGlyLeuSer----- 125
|||::: ||| ::::: |||
Db 6665 GATGATAGTTGACATTTAAGAAGACATTCAAAAGACACAAGTGTGACCAAGCGAT 6724
QY 126 -----ValLysTyrLeuArg 130
|||::: ||| ::::: |||
Db 6725 AGTTTACATGACATATTGCAAAATTTAGCTGTAGCCCTGCTATTAAAAAAGATATTTA 6784
QY 131 LysIleValLysValIleThrAsp-----LysValGluGluGln 142
|||::: ||| ::::: |||
Db 6785 CAGACTGTAAAGTTTGTGATGAATGTGTCAAAGTATGCGCGCTTAAGCCAGAAAT 6844
QY 143 -----LeuLysGlyIleThrAspGluMet 150
|||::: ||| ::::: |||
Db 6845 ATCGTTATGAAATGACAGCTGAAATACACAACTCAAAAAGGCCGCAAAAATTTCCGA 6904
QY 151 GluArgLeuArgLysAlaGlnGluValCysGlnGluLeuLys-----LysGlu 167
|||::: ||| ::::: |||
```

```
Db 6905 GACGCTATGAAACGAATTCGAAGAAGGTATCAAGAAATTAAGAAATTCATTTTAAAGAG 6964
QY 168 AsnAlaAspGluAsn---GlnLeuCysIleValGluProPheTyrSerAsnGluTyr 186
|||::: ||| ::::: |||
Db 6965 CATCTGTTGAAATACTCATTTG-----CAAAATGAAAAG 7000
QY 187 PheLeuIleValTyr-----AspValPheLysAspValArgMetValPhe 201
|||::: ||| ::::: |||
Db 7001 CTCATCTCTCTATATCTCCAAATGGAAGAGACATGATGTGACCAAGATTAAGATATT 7060
QY 202 AlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnThrMetThrProArg 221
Db 7060 ----- 7060
QY 222 HisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlu 241
|||::: ||| ::::: |||
Db 7061 -----AATCGTTTAAAGTGAT 7075
QY 242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaIleValSerMetGlnGly 261
|||::: ||| ::::: |||
Db 7076 TATGATGTCGATCAC-----ATGTCCACAAAGT 7105
QY 262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 281
|||::: ||| ::::: |||
Db 7106 TTCCTTAAAGCAGAT-----TCAATAGCAATTAAGCTC 7138
QY 282 ThrSerThrGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArg 301
|||::: ||| ::::: |||
Db 7139 TTACGCGTTCTGTATATAAAATCGTGTAAATCGGATTAACCTTCAAGTGAAGAGTGC 7198
QY 302 GlyIleLysGlnGlyIleThrPlysGluAlaMetSerAlaAspGlnIleThrArgIleLys 321
|||::: ||| ::::: |||
Db 7199 AAAAAGATGAAAACATTATGACACCAACTCTTAAGCCCAAGTTAATCACTCAAGCTAAG 7258
QY 322 TyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTyrLysAsnSerIleGlyMetAsnArg 341
|||::: ||| ::::: |||
Db 7259 TTT-----GATAATTTTAAGCAAAAGCTGAACGTGGA----- 7288
QY 342 GlyLeuAlaArgLeuAspValIleGly-----ArgLysArgAlaGluGluValGala 358
|||::: ||| ::::: |||
Db 7289 GGTGTGATGAACTGTATTAAGCTGTTTATCAAAAGCCAAATGCTGTAAACTGCCCA 7348
QY 359 PheAlaAspThrIle-----ArgLysAsnGlyLys----- 368
|||::: ||| ::::: |||
Db 7349 ATCACRAGCATGTGGCACAAATTTTGGATAGTCGATGATGATGATGATGATGATGAT 7408
QY 368 ----- 368
Db 7409 GATAAACTTATTCGAGAGTTAAAGTGATTAACCTTAATAATCTAATTAAGTTCTGACTTC 7468
QY 368 ----- 368
Db 7469 CGAAAAGATTTCCAAATTCATTAAGTAGTACGTAGATTAACATTAACATCATGCCATGAT 7528
QY 368 ----- 368
Db 7529 GCGTATCTAATGACCGTCGTGGAAGCTCTTGTATTAAGAAATATCCAAAATCTGAATGC 7588
QY 369 SerAlaValTyrGly-----AspValLeuSerSerLeuGluLysAlaTyr 383
|||::: ||| ::::: |||
Db 7589 GAGTTTGTCTATGTGATTTAAGTTTATGATGATTCGTAATTAATGATGCTAAGTGTGAG 7648
QY 384 LysGluGluLysAlaLysAlaAsnArgGluMetThrTyrLeuSerGluIleThrLeu-----PheGly 402
|||::: ||| ::::: |||
Db 7649 CAAGAAATTAGCCAAAGCAAGCCGCAAAATATTCTTTACTCTAATATCAAGAACTCTTTC 7708
QY 403 GlyThrGluVal-----ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAs 421
|||::: ||| ::::: |||
Db 7709 AAAAAGAAATTTACACTGCAAAATGAGAGATTCGCAAAAGCCCTCTA-ATCAAACTTA 7767
QY 421 PAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyrLysAspTyrLeuProSerIle 441
|||::: ||| ::::: |||
Db 7768 TGGGGAACCTGGA-----GAAATTGCTGTGGATTAAGGCGAGATTTT-----GCCAC 7815
```

REFERENCE	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM
441	unspRlysValIleuPro	AlaMetLeuSplIleValArgGlyGly	456			
7816	AGTGGCAGAAAGTATTTGTCCATTCAGCCCAAGTCAATATGTGCAGAAACAGAAAGTACAGAC	7875				
457	-----IlePro-----	-AlaAspLysLeuProAsp11	465			
7876	AGCGGATTCCTCCAGAGAGTCAATTTTACCAAAAAAGAAATATGGAGCAAGCTT	7927				
465	ePheLysAsnValIleAspLysSerValLysGlySerPheHisAla	485				
7928	-----ATTGCGTCGTAAAAAAGACTGGATCCAAAAAATATGTGCT	7969				
485	1PheAspLysSerValValProTyrSerAspLysPheHisAla	499				
7970	-TTTATAGTCCCAACGGTAGCTTATTCAGTCTAGTGGTGTCTAAGGTGGAAAAAGGCA	8028				
500	-----MetLeuLysSer-----	-MetAspLysGlyLys	508			
8029	ATCGAAGAAAGTAAATCCGTTAAGAGTTACTAGGATCACATTTATGGAAGAAGATTC	8088				
508	sPheLysAla-----	-LysAlaIleGlyLysAs	516			
8089	CTTTGAAAAAAATCCGATTGACTTTTATAGAACCTTAAGCATATTAAGAAAGTTAAAAAGA	8148				
516	ProAlaValGlyLeuSerLysSerValIleAlaIleAlaArgAlaIleGlyAlaAspAl	536				
8149	CTTATCACTTAACACTACCTAA-----	8170				
536	aMetAlaAsnAlaTyrAlaIleGlyLysGlyLysArgLeuPhePheAlaGlyLeuArgGly	556				
8171	---TATAGTCTTTTGTGAGTTGAAAAAGGTCGTAAACGAGATCGGTAGCCGGAGA	8226				
556	uMetTyrProGlyArg-----	-AlaLeuProSerAsp--	573			
8227	ATTACAAAAAGAAATGACCTGGCTGCTGCCAAGCAAAATGTGCAATTTTATATTATAGC	8286				
573	tSer--TyrGlySerIleLysGlyTyrGlyLeuProLysAspGly-----	588				
8287	TAGTCATTATGAAAGATTTGAAGGCT--AGTCCAGACAGATTACGAACAAAAACAAATTGTT	8343				
588	pTyrAsnTyrHisThrThrGlyLysGlyValLeuGlyLysGlnAspProLysSerAspGly	608				
8344	TGTGGAGCAGCAT-----AACCATTTATTTAGATGATGATTATTTGAGCAAAATCAGTGA	8394				
608	uPheLysAlaGln-----	612				
8395	ATTTCCTAAGCGTGTATTATTTAGCAGATGCCAATTTAGATATAAGTCTTATGCTATPAA	8454				
613	-----GlnAsnIleLeuAspLeuPheArgThr	621				
8455	CAAACATGAGAGCAAAACCAATACCTGTAAGCAACAGCAACAAATATATTATTCATTATTTACGTT	8514				
621	LysAsnTyrGly-----	-Arg	626			
8515	GACGAGATCTTGAGAGCTCCCGCTCTTTAAATATTTGATACCAACAAATGATGATGTAAG	8574				
626	gTyrAlaGlyAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAspArgIleThrGly	645				
8575	ATATACGCTACAAAGAAAGTTTATGATGACCACTGTTATCCATCAATCATCATCTACTGCT	8632				

TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS
Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Autay, M., Huber, R., Feldman, R. A., Short, J. M., Olsen, G. J., and Swanson, R. V.	Nature	392 (6674), 353-358 (1998)	The complete genome of the hyperthermophilic bacterium Aquifex aeolicus	
2 (bases 1 to 15569)				
Decker, G., Warren, P. V., Gaasterland, T., Young, W. G., Lenox, A. L., Graham, D. E., Overbeek, R., Snead, M. A., Keller, M., Autay, M., Huber, R., Feldman, R. A., Short, J. M., Olson, G. J., and Swanson, R. V.	Submitted		Direct Submission	
Submitted (25-JUL-1997)	Diversa Corporation, Genomics, San Diego, CA 92121			
COMMENT	Hypothetical indicates no similarity to known proteins			
FEATURES	Hypothetical indicates similarity to a protein of unknown function.			
SOURCE	Location/Qualifiers			
gene	1..15569			
gene	/organism="Aquifex aeolicus"			
gene	/strain="VPS"			
gene	/db_xref="taxon:63363"			
gene	complement(199..1497)			
gene	/gene="pura"			
gene	/note="ag_1290"			
gene	complement(199..1497)			
gene	/gene="pura"			
gene	/codon_start=1			
gene	/transl_table=1			
gene	/product="adenylosuccinate synthetase"			
gene	/protein_id="AAC07286.1"			
gene	/db_xref="GI:2983724"			
gene	/translation="MEKILILGAQMGDEGKGIYDLILSEHDDIVRYOGSGNAGHTVAVNSKRFILHLPTGILHEDHVKGVYAQGVYDLVLEHKEVKNLEKGIYVKRLTISDRALHWLPKRLDLSLEKKGIGITLRIGIPATMYKGGKIRISDLDERKFTYLLLELNDLVPKNCLEKVFCEKEDLDLINOIYEQRLRFEFEKENVDDLFPYOKGSYLFESAGOTGLDDVMGTGYPYVTSNSMAGLSMGTGMPKPYSDAFELGVAKAYTTEGAPPEPTELGESEKRELGEYSGTGRPRCGMDPLVALKAAVONLDGCVITKLVDLTLDLFDKVCVAVLELDEGVITDIPASYSILIRVVPVYTKLGMKSKSKAKADVSELKEMALDYKVFIEYTGVPVYMLSTGCRDEYITWKLILTRSGYS"			
gene	complement(1526..2062)			
gene	/gene="ag_1292"			
gene	complement(1526..2062)			
gene	/gene="ag_1293"			
gene	/codon_start=1			
gene	/transl_table=1			
gene	/product="hypothetical protein"			
gene	/protein_id="AAC07294.1"			
gene	/db_xref="GI:2983732"			
gene	/translation="MKIILGEGVGTLLVKKIVLERLGRAGISFMTEEVDPEPKRRTGFEITITTEGKKKIFSSKFFTSKLVSGYGVNVPFEELAILERKATREKKRRRYIIIDIEGKMELESKRFPDVLVQIMHDPNVNVAITPIRDVPLVKEITRLPGAVLIEIPENRDVILIEDILSLER"			
gene	2131..4734			
gene	/gene="alas"			
gene	/note="ag_1293"			
gene	2131..4734			
gene	/gene="alas"			
gene	/codon_start=1			
gene	/transl_table=1			
gene	/product="alanyl-tRNA synthetase"			
gene	/protein_id="AAC07289.1"			
gene	/db_xref="GI:2983727"			
gene	/translation="MSLSAHEIRLELFSFEKKGHTRVKSAPLVPENDPTLLFVNAGMVPFNVLGLEKRPYKRRATSCOKLRVSGKINDLEQGYVYSRHPTFEMILGNFSFGDYFKKAIETAYEMFEVTLKPEKLYSVYKDEEATYIAMEHISIPSEIRIWLREEDNIDFQWQDVGPGSGSEIYVNGEEYEGSERLYETINLVMQVNRDENVLPRLPHNIDITGMLEIRASLYLQSKNSNFEDITFIPLIOGEEBVGKKGKYEKETVALYIADHLARKTMSLISGVATPSNEGRGVIRILIRKMRFGYKGIENPFIKGVLDVLDIMKPEPTELISREVPKGVKVEEKRFLTKLAKAMEYIOEVLOKALEGRKTLQSKVEFTADYTGFPVDLIDIEAREKGLIDEGFOCELEORERARKHRYKAKPYSHLKEGKTSFAFYIEMHESQVGLVKGGLVSELKEGEEGVVLKETPIYEGGQIGDAGIIESSDKALFKEDYQKPREGIIYHIGVLTGKTLGTLGVHARVDKERMDIMRNTATHLHAAALNVEGHERQAGSLVADKYLRFDTTHSALTDEELKRYEELVNEKIRNTPLVNMAADVDELKGAALIFEERKGERVATVSCGFSLELGGTHVSTGDIIGFKIISSESS			

```

gene
/4739..5209
/gene="aq_1295"
/4739..5209
/gene="aq_1295"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC07293.1"
/db_xref="GI:2983731"
/translation="MKKEETRTGILKDVSIIRLVSPSPLLPGGFPVCEYGIPIRYLX
IDENYKLEISYVEKLNLSRVCNIEQVYDVKVEELRYALPEGVFTIAYKANGITLR
ELSYLEIYTSVGIIIDEDRELDOEPAPVAKENRTIIGTIREVILLEIMYS"
5306..7738
/gene="clpc"
/note="aq_1296"
5306..7738
/gene="clpc"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease"
/protein_id="AAC07290.1"
/db_xref="GI:2983728"
/translation="MFEKTERAROVILRARERALELGHSLGSEHLLALIDEDIP
VVLVSKLGVSPKIRKIMGOITIKGSHSGEVLAPDPKRYIEFAVEAPRLHIOFGP
EHLILVYRKETDGLGRILKSFGLDEYVAREVLQILGELPQESQRYSTPMIDRS
RDLTOMAREGKLDLVIGREKEIEIRVIOILVRRKNPVLGLDGPVKTALVEGLAORI
ANKEVEPLINKRVAALDLAALVAGTIRGOEFERKNIILEKANNVLLFIDELHT
LVGSAEGSIDASNMKLPAALARGEIOVIGATVDEKRYIEKDGLEREFQVYVQ
PSPEDITOLILYGLKPKPEEFHKVETPEALEKAVELSVKITYDPNRIDPKRIDMDAG
AUKIKATOLPELKEIEEKIKLEKEKEKAYVEQDYERAAKIRDEELRIKRAETIK
AKKEMEKNRPKRVDEDAEVAVRWGTPVKRVHSDAKLLHIEBELKRVAGDE
ALAAVAKAIRSRVLGKGRPIGVEMFLGPTVGRTETAKALAEYLFCTEDALIRD
MSEYMEKHTVSRILGAPRGVYEGGOLTEAVRRPYSVLFDIEKAPDVFNFL
QIPDDGLTDLGRTVFSNTIIMTSLNGARLITGTMGFEKKOGLNFEEMKRVNL
DOYKRTFSPFLNRDEIIVFRLEKKIDQILQILQIKETIKDMNVVVKLHKSIV
NMIIKEKPEFGARSRLRQLORHVEDLAEELIKGTAKGVEEVEIRVRNDRKYLKPY
KKKEKVALIEK"
7776..10112
/gene="omp"
/note="aq_1300"
7776..10112
/gene="omp"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AAC07292.1"
/db_xref="GI:2983730"
/translation="MKKEPLTALFLGVANAARIEKIIKGNKYIDELIRGLLSVRK
GSEFSLERIKRDLRKILRGLFTKVEYIEEDOKNGRPEVILVEDLPVLYKIEFBN
EEFSDDLKELYLIELELGEVEEETIEGTSSPAIEERLAIVKOLGLILFOELE
SLRRRIEYVYKRGHFGVTKEVYLIPKKGSKLVPIHEKKITSDIEEGKMHSD
RLKUGLMELEKPNIFLRLHAFSEELKEDVKKIEEYNEEFLSEYVEYKESD
ARKVIVKRGPRYKLEIKIEGNTLFAVSELDNIIKKRKEGRFRFRVRIKIRN
IRKVAIEGFLNTEVERVNVPEKKEVSLKLTIGGKPYVKIKITIKYERDVIY
RRMRVQENELAKKGIERSKTRIMNIGYEDVQIEPVRPDMWDLVAKIRRFQO
FSVGLSYNEVYGLAGFLELKNFREGTDLAGISVSGLSYRNNAISYRWKLRKRV
DLDLSAFNRILEYDYVERTGFSVALSKELSEYWRASIGTSIORVYSDIDEASTY
VKDQAGRDRSKIFETLIRDTDYLLPTGSLFERNVSGVLLGDEKFFYFEVGG
AKYESPVEYDGTIILSLKGEVVEGGKKVPIDEEFVGGFSGISRGYVYAGAPVD
PMTLDLIGAKKIVTISLELMYPVFKMLYFAGFEDYGLGADKMSDRKLSFRGTYGIG
VRITTPAPRIRIDMARKTVPEGDDTNRSHFVLGTFP"
complement(10138..10350)
/gene="cspc"
/note="aq_1303a"
complement(10138..10350)
/gene="cspc"
/codon_start=1
/transl_table=11
/product="cold shock protein"

```

```

gene
/protein_id="AAC07291.1"
/db_xref="GI:2983729"
/translation="MSFRGYKRFKFSODKGGFITTREDTNADVVFHTDIOMEGFKTLQ
KGGKVEFDVEDTKGPPAKNRVLGE"
complement(10454..11170)
/gene="hisa"
/note="aq_1303"
complement(10454..11170)
/gene="hisa"
/codon_start=1
/transl_table=11
/product="phosphoribosylformimino-5-aminoimidazole
carboxamide ribotide isomerase"
/protein_id="AAC07284.1"
/db_xref="GI:2983722"
/translation="MNLEKFTIPALDMEGKAVRLYKGFNKKVYSERPWEIKNFS
DLGFKRLHVVLDLEAGGKRPKNLEVRIRKIRENFEGEVGGIRSYEVALKALDEGID
FVIGITLAVYKKEEFLKILENPNRVLIAIDSKQGVVAILGGMKEETAAPSPEEPAKYE
NYPIMGYLVTVIEODGSLGSDVDEPYKEIKKHVKKPIVASGVSLIEDIKKIYIYEG
VYVGRKAIYEGRTLEDLONP"
complement(11167..11922)
/gene="pyrD8"
/note="aq_1305"
complement(11167..11922)
/gene="pyrD8"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase electron transfer
subunit"
/protein_id="AAC07287.1"
/db_xref="GI:2983725"
/translation="MLAEINAVENTYISGNLYRLTIRIPDKILKOIEPGHAMIKP
SPDYDMGRAPAVADIEBKAFYDYDGKGYILSKRGRSLTPTPLDKRLSY
EGKHLLIGGAGVGLAGLTLAKLRDMGKRVFATYGRSKREHLGMYEMKEEFPYEL
FTNDGSAGRGVETELIKESTDWVAVACGPKPLTIEMKTGKRVYVSLERRMAG
MGVCLCIVYKTRGKFRKRVCEYGPVAPMEVEYL"
11990..13300
/gene="sucC1"
/note="aq_1306"
11990..13300
/gene="sucC1"
/codon_start=1
/transl_table=11
/product="succinyl-CoA ligase beta subunit"

```

Alignment Scores:

Pred. No.:	Score:	Length:
3.04	144.50	15569
Percent Similarity:	34.18%	Matches: 111
Best local Similarity:	20.68%	Conservative: 111
Query Match:	3.89%	Mismatches: 267
DB:	1	Indels: 277
		Gaps: 40

```

US-10-008-355-2 (1-712) x AE000733 (1-15569)
QY 9 LeuLeuGlyAlaAlaLeu-----LeuLeuGly-----AlaSerGlyValAla 22
||||| ||||||| :|||
DB 3820 CTGCTTACGCCGCCGACCTCAGGAACGTTCTCGGACGACGACGTAACAGAGCGTCTCTC 3879
QY 23 LysAlaAspIysGlyMetIryLeu-----LeuSngluLeuAsnGluAsn 38
||||| :||| :|||
DB 3880 GTCGCGGACGAAGTATTGAGGTTTGACTTTCACACATTTTCAGCTCTAACGGAAGAGAA 3939
QY 39 LeuAspIryMetAlaGluLeu-----GlyPheMetLeuProLeuAspSer 53
||||| :||| :|||
DB 3940 CTCGAAGAGGGGGAAGAACTGTGAACGAAAAATAAGGCAAAACCTCCCGTGAATGTA 3999
QY 54 Leu---TyrSerPheAspIysProSerIleAlaAsnAlaValIlePheGlyGly 72
::: :||| :|||
DB 4000 ATGGAATGCGATACGATGAAGACATAAAAGCGGTGCGCATAGCATATTGAGGAAAG 4059
QY 73 -----CysThrGly 75
||| |||

```

Db 4060 TACGGGAAAGGTGAGCGTAATATCTCGCGAGTCTTCAAGAACTCGCGGTGA 4119
 QY 76 ILeThrValSerAspGlnGlyLeuIlePheThrAsnHisCysGlyTyrglyAlaIle 95
 Db 4120 ACCGACGTTCCGCACTGAGACATN -----GGCTACTTCAGATA 4161
 QY 96 GlnSerGlnSerThrVal ----- 101
 Db 4162 ATATCCGAGAGCTCCGCGCGCGGTGAGAGAAATAGTGACACAAACGGAGGTG 4221
 QY 102 -----AspHisAspTyLeuArgAspGlyPheValIleSerArgThr 114
 Db 4222 TCCGTAAACAACCCCTTCAAGAACACAGACGCTCAAGAACTTCGACGCTCGC 4281
 QY 115 MetGlyGlnGluIleuProIleProGlyLeuSerValIleTyLeuArgIleValIys 134
 Db 4282 GTGGCGCAGAG -----GAAAGTAATCCAG 4305
 QY 135 ValThrAspLysValGlnGlyGlnLeuLysGlyIleThrAspGlnMetGluArgLeuArg 154
 Db 4306 AAATTCGAAGAACTTAAGAAATCAAGAACAGAGAAAGAAATCAAGAACTGAAA 4365
 QY 155 LysAlaGlnIleVal -----CysGlnGlnLeuAlaLysLysGlnAsnAlaAsp 170
 Db 4366 -----CAGGAAGCTCTGAAAGCTCCAGATTAAGGAAAGTTGTAAAGGAAAGAAAGCTTGA 4419
 QY 171 GlnAsnGlnLeuLysIleValGlnProPheTySerAsnAsnGlnIleTyPheLeuIleVal 190
 Db 4420 GAT -----TTCAACACTCCAC 4434
 QY 191 TyrAspValAlaPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPhe 210
 Db 4435 TACGCGCTATTGAAAGAGTTGAA -----CCGAAAGAACTCAGAAACCTC 4479
 QY 211 GlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPheArg 230
 Db 4479 ----- 4479
 QY 231 ValTyraIleGlyAlaAspAsnArgProAlaGlnIleTySerLysAspAsnLysProTyLys 250
 Db 4480 -----GCGGATATGCTGAGGCAAGGCAAGCAAGAAAGAAC ----- 4512
 QY 251 ProValTyThrPheAlaAlaValSerMetGlnGlyTyTyLysAlaAspTyraIleMetThr 270
 Db 4513 -----GTTGTGTTTATAGCTCCAGAAAGGAGATTAAGATAAATTCGT 4557
 QY 271 IleGly-PhePro -----GlySerThr 277
 Db 4558 ATGGAGATTTCGAAGAGATATCGACAAAGCTGAACGCCAAGAACTAATTAAGGAGGTC 4617
 QY 277 TrsAspArgTyraLeu -----ThrSerTrpGlyValGlnAspArg 289
 Db 4618 GGAAGGTAATCTTAAGGTGAGAGAGAGAGAAAGGACAGACTCGCTCAGGTGAGAGAAAG 4677
 QY 289 GlnIleGlnAsnGlnAsnPro -----ArgIleGlnValArgGlyIle 303
 Db 4678 GCTCCGTCACA -----AATTCGCCGGAAGCGTTAATTAAGGAAATCTTAATGCGGTG 4732
 QY 303 elysGlnGlyIleIleTyPheGlnIleMetSerAla -----AspGlnIleAlaThr 318
 Db 4733 -----AAAGGTGTGATGAGAGAAAGACAGACAGACGCGGTATCGACTCAAAAGCTGTGAT 4788
 QY 318 ArgIleLysTyraIleSer -----LysTyraIle 327
 Db 4789 AAGGCTGTGAGAGAGAGCTCCCTCTCCGCGGAGGCGTTCGTTACTGTAGTAGAGG 4848
 QY 327 ArgIleSerAlaAsnTyraIleTrpLys -----AsnSerIleGlyMetAsnArgIleVal 344
 Db 4849 AATCCCGGTAAAGTAACTGAGATAGACGAAGAAAGCAAGTAAGAACTGAGATTCCTGGA 4908
 QY 344 aArgLeuAspValIleGlyTyrglyAsnArgAlaGlnGluArgAlaPheAlaAspTrpIleArg 364
 Db 4909 AAAGCTGAATAATTTATCGAGGAGGTGTGTGAATAACATAGAACAAATTGTTGACAAAGTTGT 4968

QY 364 gLysAsnGlyLysSerAlaValTyrglyAspValLeuSerSerLeuGlnLysAlaTyrgly 384
 Db 4969 AGAAGAGTTGAGGTAGCCCTCGCCAGGGGTGTTTC -----ACTATAGCGAAGTTTACCG 5025
 QY 384 sGlnGlyAlaLysAlaAsnArgGlnMetThrTyLeuSerGlnIlePheGlyIleVal 404
 Db 5026 AAAGGTAATTAACCTC -----AGAAAGCTCTTCACTTGAAGAG ----- 5065
 QY 404 rGluValAlaArgPheAlaGlnPheAlaAsnAlaLeuAlaIleAsnProAspAlaHisAl 424
 Db 5066 -----TACACTTCCTCGGTAAAGAAATTAATTAAGACAGAGAC ----- 5104
 QY 424 agIleLeuLysSerLeuAspAspLysTyraIleAspTyraIleProSerLeuAspArgly 444
 Db 5105 -----AGGCACTGTGATCAGGAATGCCCGCTTACGTTAAGAAACAGGGAAC 5154
 QY 444 sValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAs 464
 Db 5155 GATCTCGCGCACTATCAGGAAAGTCACTCGCTTGAGCTCATGTACTTCTTAAAGAA 5214
 QY 464 pIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysTyraIleAspThr 484
 Db 5215 CATTAATTAAG -----ATATATACACTGTTTGTCAATAGCCTTGAAAGAACTTATGTTT 5271
 QY 484 eValPheAspLysSerValValProTySerAspLysPheHisAlaMetLeuLysSerLe 504
 Db 5272 TTTATTTTATTAATAAAGCTGAGAGAGTAAAGACATGTT ----- 5311
 QY 504 tAspLysGlnLysPheAlaLysAlaIleGlnLysAspProAlaValGlnLeuSerLysSe 524
 Db 5312 -----GAGAAAGTTC -----ACAGAAAGGCAAGACA 5337
 QY 524 rValIleAlaAlaArgAlaIleGlnIleAlaAspAlaMetAlaAsnAlaTyraIleGln 544
 Db 5338 AGTTATCTTGAGGCGCAGG -----GAAAGAGCCTCGAGCTCGACACGCTACTT 5388
 QY 544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGlnMetTyraProGlyArgAlaLeuPr 564
 Db 5389 GGAAGATGAACATATTTGCTGCTCATTAAGATGAAGAGCAATACCGCTCTGCTGC 5447
 QY 564 oSerAspAlaAsnPheThrMetArgMetSerTyrglySerIleLysGlyTyrgly --Pr 583
 Db 5448 TCAGCAAGCTGGGTGT -----CTCTCTGAGAAAGATTAAGAAAGGCAATATGAGGCT 5498
 QY 583 oGlnAsp -----GlyAlaTrpTyraAsnTyraHisThrThrGlyLysGly -- 597
 Db 5499 AATTAATTAAGGAGACCACTCGGAGAGG --TACTCTTCGACCGACGCAAAAGGGA 5557
 QY 597 ----- 597
 Db 5558 ATAGATTTGCAAGTGAAGAGGAGAAATTTCTACACACAGTTCTAGGTCCGAAACAC 5617
 QY 598 -----ValLeuGlnLysGlnAspProLysSerAspGlnPheAlaValGlnIle 613
 Db 5618 CTACTATGCAAGTGGTGTGAGAAAGACGGGCACTCGGTGAAGAAATTT -----TA 5668
 QY 613 uAsnIleLeuAspLeuPheArgThr -----LysAsnTyraGlyArgTyraIleGln 630
 Db 5669 AGGAGCTTTGGAGTGCAGCAAGTACTCCGTAAGAGAGAGACTTCACTACTCGGAGAA 5728
 QY 630 nGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleHisGlyLysSerGlyLys 650
 Db 5729 CTTCGCCCTCAAGAAAGCC -----AGAGGTATCGTA --TCCAC 5763
 QY 650 rProValPheAspLysAsnGlyArg --LeuIleGlyLeuAlaPheAspGlyAsnTrpIle 669
 Db 5764 TCCTTAACCTTGACAGGTTCTCAAGGACCTCAACCCAAAGGCAAGGAAAGCTTCA 5823
 QY 669 uAlaMetSerGlyAspIleGlnPheGluProAspLeuGlnArgThrIleSerValAspIle 689
 Db 5824 CCCAGTAATAGCGAGG -----GAAAGGAAATTAAGAGAGATTATCCAGATTAATCTCGT 5874

QY 689 earg 690
 Db 5875 ANGA 5878

RESULT 5
 OSJN00169
 LOCUS Oryza sativa chromosome 4 clone OSJNBa0043124, *** SEQUENCING IN
 DEFINITION
 PROGRESS ***, in ordered pieces.

ACCESSION AL662969
 VERSION AL662969.1 GI:17998481
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
 1 (sites)
 Han, B., Feng, Q., Huang, Y.C., Li, Y., Zhu, J.J., Zhao, Q., Hu, X.,
 Liu, Y.L., Mu, J., Yu, Z., Chen, L., Fan, D.L., Meng, Q.J., Zhang, L.,
 Lu, X.Q., Yu, S.L., Liu, X.H., Lu, T.T., Zhang, X.J., Lu, Y., Li, C.,
 Li, T., Zhang, Y., Hu, H., Jia, P.X., Qian, Y.M., Ying, K., Zhou, B.,
 Chen, Z.H., Hao, P., Zhang, L., Wu, M., Zhang, R.Q., Guan, J.P., Fu, G.,
 Wang, S.Y., Ren, S.X., Lv, G., Lin, W., Gu, W.Q., Zhu, G.F., Tu, Y.F.,
 Jia, J., Yin, H.F., Zhang, Y., Cai, Z., Chen, J., Kang, H., Chen, X.Y.,
 Shao, C.Y., Sun, Y., Hu, Q.P., Zhang, X.L., Zhang, W., Wang, L.J.,
 Ding, C.W., Sheng, H.H., Gu, J.L., Chen, S.T., Ni, L., Zhu, F.H. and
 Hong, G.F.
 Direct Submission
 Submitted (27-DEC-2001) Han Bin, National Center for Gene Research,
 Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
 CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
 bhan@ncgr.ac.cn
 Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
 clone: OSJNBa0043124.
 Web site: http://www.ncgr.ac.cn
 ----- Summary Statistics
 Assembly program: phrap

COMMENT
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1.162595
 /organism="Oryza sativa"
 /variety="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="4"
 /clone="OSJNBa0043124"
 /clone_1lb="CUG1-OSJNBa"
 BASE COUNT 44942 a 35736 c 36400 g 45514 t 3 others

FEATURES
 source
 1.162595

Alignment Scores:
 Pred. No.: 168 Length: 162595
 Score: 137.50 Matches: 147
 Percent Similarity: 36.64% Conservative: 115
 Best Local Similarity: 20.56% Mismatches: 260
 Query Match: 3.70% Indels: 195
 DB: 2 Gaps: 34

US-10-008-355-2 (1-712) x OSJN00169 (1-162595)

QY 47 PheThrLeuProLeuAspSerLeuTyrSerPheAspLysProSerIleAlaAsnAlaVal 66
 Db 108215 TTGTGTCAGCCTCTCGCTTTT-----TCACATGCAAGTGGTACA 108256

QY 67 ValIlePheGlyGlyGlyCysThrGlyIleThrValSerAspSerglnGlyLeuIlePheThr 86
 Db 108257 AGTCACCTTGATGTCAGCTTCAGCTATGTGTCATGACAA----- 108301

QY 87 AsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeu 106
 Db 108302 -----ACTGCTGCTATTCTTCAGATTCATCCCTTTTCATCAATACTCTT 108346

QY 107 ArgAspGlyPheValSerArgThrMetGlyGlnGluLeuProIleProGlyLeuSerVal 126
 Db 108347 CGTTTGTCA---CAGTCAAAGACTGTGGAGAGAG-----TGTGA 108385

QY 127 LysTyrLeuArgGlyIleValLysValThrAspLys-----ValGlnGlyGln 142
 Db 108386 AGAAGCCCTAGAGAGTGTCTTAAGTCAAGCAAGAAAGAGTTTCCATTTGAAGTTTG 108445

QY 143 LeuLysGlyIleThrAspGlnMetGluArgLeuArgLysAlaGlnGluValCysGlnGlu 162
 Db 108446 CTGAAGGTGTAGC----- 108460

QY 163 LeuAlaLysLysGlnAsnAlaAspGluAsnGlnLeuCysIleValGlnProPheTyrSer 182
 Db 108461 ATATCAGATAGGCAAAATATCTCAGCTACCGCTCAACAGCTGGATCT-TGGTATTTCT 108519

QY 183 AsnAsp-----GluTyrPheLeuIleVal-----TyrAspValPhe 194
 Db 108520 TTTAAGCTTAACTGTCTCAATTAAGTGAATGCGCTTTTATCATGTATTT 108579

QY 195 LysAsp-----ValArgMetValPheAlaProProSerSerValGlyLysPheGlyGly 212
 Db 108580 CCTGGTCATTAAGTCAATATATATCATCTTAAGTTTCTATGATTAATAAATC----- 108633

QY 213 AspThrAspAsnThrPheTyrProArgHisThrGlyAspPheSerValPheArgValTyr 232
 Db 108634 -----AGTTCTTATTTCT-----TTTAACTCAAGT 108660

QY 233 AlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProVal 252
 Db 108661 GCAGGGGGTGTAT-----CCTCATCGCTTCGGAT-----CCTCTGTT 108699

QY 253 TyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspPyrAlaMetThrIleGly 272
 Db 108700 CCGCTTGCAACAGACAGCTTCAATATCATCTGTCTTACAGATTCAGACCTCGACGCTCA 108759

QY 273 PheProGlySerThr-----AspArg 279
 Db 108760 TCTGTTCCTAGCACAATAATGCTAGTCAGCAAAATGGAGGCTCCCGCTTATGGAGTCA 108819

QY 280 TyrLeuThrSerTyrPheValGluAspArg-----IleGlu 291
 Db 108820 ATGACAACCCAGTGTGGCAGCAGAGAGCATCAAGATCACCGTACTGGTATATATCA 108879

QY 292 AsnGluAsnAsnProArgIleGluVal-----ArgGlyIleLysGlnGly 307
 Db 108880 TCTGATCTCCATGCAAGACTTATACGTCTTTCCAAAGAGGTCTTTGGAGAGGCCA--- 108936

QY 308 TyrPheGluAlaMetSerSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAla 327
 Db 108937 ---CAAGAAGGAGGCGGCATGATGATGAGGAGTATATCCGTCACACTAGGAATTCGCC 108993

QY 328 GlnSerAlaAsnTyr-----TrpLysAsnSerIle----- 337
 Db 108994 CAACACTCAAAACTATGTTCATCATCCCTTACAGAGATGCCATCCACAGATTCACATTAAT 109053

QY 338 -----GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArg 353
 Db 109054 ATATCATGTTCCAAATTTCCAGAGACCTCTTCAAGAGACGACAGATGTCAGAGGCTTCT 109113

QY 354 AlaGlnGluArg---AlaPheAlaAspTyrPheArgLysAsnGlyLysSerAlaValTyr 372
 Db 109114 GCCAGATCAGCAGCAGCTTTGACATGACGACCAAGTACAGTACAGCTACGCTGTGTTAT 109173

QY 373 GlyAspValLeuSerSerLeuGlnGluLysAlaTyrLysGlnGlyAlaLysAlaAsnArgGlu 392
 Db 109174 ACGGATGACACTGCGCTTCACTAAGTGAATGATGCTTCTTGAAGGAT----- 109215

QY 393 MetThrTyrLeuSerGlnThrLeuPheGlyGlyThrGlnValValArgPheAlaGlnPhe 412

```

Db 109216 -----CTCAGCCCTAGTTCGGATG--GGTAGGAGAGCTCAGCTTTGAATTAT 109265
Oy 413 AlaasnaIaLeuAlaThraSnpProAspIaHsAlaGly-IleLeuLysSerLeuAspAs 432
Db 109266 TCGGAATTTGCGTACACAGGCGATATTCAGAAATTCACAGAAATTTGGA 109325
Oy 432 PLYsTYrIaSPYrIaPheProSerLeuAsp-----ArgLysValIleProAlaMe 449
Db 109326 GGTGATGAAGCGTTTTCGTCATGATGACCCCTCATATGATTAAGTTGCAAGCAGC 109385
Oy 449 tLeuAspIleValIArgArgArgIleProAlaAspLys----- 461
Db 109386 CTTTCTACTCTCTTCAGAACTATTATTCAGCTGCAGAGAACCATTTGAGATTATGTTGA 109445
Oy 462 -----LeuProAspIlePheLysAsnValIleAspLysPheLysGlyAspThrLys 479
Db 109446 AAGAAATTTGCGCATATGTTTTTCCCGGCTTATGATCCAAAGAAATG-----GTAAA 109499
Oy 479 sLYsTYrIaAspPheValIleAspLysSerValValProTyrSerAspLys-----Ph 497
Db 109500 AAAACCATGCTCATCAACCTTGATGTTGTTGGAGAAACATTCGAAATTCATATGTTGCT 109559
Oy 497 eHIsAlaMetLeuLysSerMetAspLysGlyLysPheAlaLysAlaIleGlyLysAspPrr 517
Db 109560 ACCTCGCTAGTGCCTCATAGATGAACAGAGTCCCAAGGCA-----AACTTGCC 109613
Oy 517 oAlaValGluLeuSerLysSerValIleAlaAlaIaArgAlaIleGluLysAlaLeu 537
Db 109614 TGTCTTGAAGTTGCGCAAAATAATGCTTC--AGCAAGTACACTGTATGATTCGAAGTTA 109670
Oy 537 tAlaAsnaIaTYrAlaIleGlu----- 544
Db 109671 TAGTAACAGTGGGCTTCCTAAACGTGGCTTTCANAAATTCGCACTTATGTTATCATGAAA 109730
Oy 545 -----LysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetLysTyrProGlyArgAl 562
Db 109731 AAATGCAAAAGTGAAGGAGGATCCATTTCTGATCATATCTGTTATTCATCATTTGA 109790
Oy 562 aLeuProSerAspAlaAsnPheThrMetArgMetSerLys----- 575
Db 109791 TTCACACAGAGCTGAATTTTATTCCTAACTATCTGTTAGAGGCAAACTTTGAG 109850
Oy 576 -GlySerIleLysGlyTyrGluProGln--AspGlyAlaTYrPyrAsnTyrHisThrTh 594
Db 109851 GCGTCGACTGAAGCAATGACACGCTCGATGAACTGATTTAGTGAAGTCTTACAGAG 109910
Oy 594 rGlyLysGlyValIleuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAs 614
Db 109911 CAAAGAA-----GATCGTCCCGTCCAAATCTTATGATCAAGCA----- 109950
Oy 614 nIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuH1 634
Db 109951 -----GATTATGCAACTTCTCCGAAGATG----- 109977
Oy 634 sIleAlaPheLeuSerAsnAsnAspIleThrGlyLysAsnSerGlySerProValPheAs 654
Db 109978 -TACCGGCTTGCTCAAGAAAGCTATCCATTTGCGGCGATTTCTAGTCCCTTGA 110036
Oy 654 pLYsAsnGlyArgLeuIleGlyLeuAlaPheAspLysAsnTrrp 668
Db 110037 TGCTGAAGT-----GGGAAATGG 110055

```

```

REFERENCE 1 (bases 1 to 4004)
AUTHORS Kimura,T. and Horikoshi,K.
TITLE The nucleotide sequence of an alpha-amyLase gene from an
JOURNAL alkalophilic Micrococcus sp
FEATUERS Microbiol. Lett. 71, 35-42 (1990)
source location/Qualifiers
1..4004
/organism="Micrococcus sp."
/strain="207"
/db_xref="taxon:1271"
165..189
/rept_type=INVERTED
265..272
RBS 282..3596
CDS
repeat_unit
RBS
CDS
/transl_table=1
/product="alpha-amyLase"
/protein_id="CAA39321.1"
/db_xref="GI:296763"
/translation="MSKKMKOKFLMVLALLVIGDAVCSSATGATSTPAADREVR
LSTIVRINDDMIFGRHRRRLMKSTLSRMKGVRSRVAVAPRREGIWIQFAOL
KLDSFPRKINGSLIOVNKDSLQAEVTSIGTEIRIVPDSAPVIDNGENMFTYRDE
LFADADAMNTEKVELDFDGETHEWTEPEPEREVEFLANDIPTEBHPYTLVRAQVTT
EVDIPYNTVGVDSITLVNQNADLAVSGTVAPEALIDNONAVSDIEGITEDVATISKIY
ADTSLGSAOLEIDPTLKEISLAVDDIADAGRAIPLTVDSGYNLKRGLAVAEVAT
ROSYGEGDFPDMDSVITYEMLTDRFDPOAFANDOVEYENILAHYTRILFREGYSKVEAK
LDYIDELGVNTISPVENYIKDVRERYESPEYCYCHGYMANFGLNDFLPTMEEF
HDILDGADHDMKIMVDVNVNHTGYLKEIDGSTVNTPEPATPSDADARFSDLROCA
DVGTDEVVGEIAGLPDFTEDPNVRKQIDMOTDIMEKATTEGNTIDYRVTYKIVH
EDATWMOFKNALTEKMEPFKIGEWAKVDNLTGLYLETOTMSLIDPFEPKERSRV
NGSLAEANASLTARNAKLDNTATLGOFLSGHDEGFLHSLADRGKLOVAATLQATAK
GOPYITYGSEELGOTGANNPYOYDRYDFANDVEYENILAHYTRILFREGYSKVEAK
GERTVGSGDKDFLEFSRDYOKYVYVGLNVAESEAVTLTVDSGYNLKRGLAVAEVAT
YTRATAGVNNLTLPKADPTGYTLTVEEGNTTYCAKDNGEVYVELVPENNIRIHKRED
NVYKNYGAMLMNDVAPSSANVPVATVEKTDSTGATIDVPLDGAANNISFLVMDITA
GDAGKDGDDGFTLISPOANEIWKOSDSVYVEPYDLPANTVRIHYTREAADYDTA
GIWNMGDVASPDGMPTGAVNFDGTDYGAIPDKLEAKKEIGFLVNLNKEADYDAG
DKSFNLIDKFNHLMWIKOGDDTVYVSPYDEVASGLSNAEVSENNTLILRFNSTDGTSE
KIVAGLIAIDPADSAAL"
BASE COUNT 1239 a 724 c 1028 g 1013 t
ORIGIN
Alignment Scores:
Pred. No.: 3.45 Length: 4004
Score: 133.50 Matches: 138
Percent Similarity: 34.94% Conservative: 94
Best Local Similarity: 20.78% Mismatches: 259
Query Match: 3.59% Indels: 173
DB: 1 Gaps: 35
US-10-008-355-2 (1-712) x MSP207AAM (1-4004)
Oy 141 GlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuAlaGlnGluValCys 160
Db 1431 GCGCATTTTAAAGCATTCACAGCAAGCTGGATTAATTA----- 1469
Oy 161 GlnGluLeuAlaLysGlyAsnAlaAspGluAsnGlnLeuCysIleValGluProPhe 180
Db 1470 -----GATGACCTTGCGGTTTATATACAAATTGGATT--AGTCCGGTT 1508
Oy 181 TyrSerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetVal 200
Db 1509 GTGGAAGAACATTAAAGTAC-----GATGCTCGTTATATAC 1541
Oy 201 PheAlaProProSerSerValGlyLysPheGlyAspThrAspAsnTrrp-----Met 218
Db 1542 GAAGCGTCAGAACCGTATTAATCATCAGCGTACTGGCAAAATAATTTTGGAGAGTTA 1601
Oy 219 TrpProArgHisThrGlnAspPheSerValPheArg--ValTYrAlaGlyAlaAspAsn 237
Db 1602 AATCGG---CATTTTGGCAGCATGGAAGAGTTTCATGACTTGATGACGGGGCACACAGAC 1658

```


TITLE
JOURNAL
Direct Submission
Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES
source
1.10292
location/Qualifiers
/organism="Streptococcus pneumoniae R6"
/strain="R6"
/db_xref="taxon:171101"
147. .341
/rpt_family="BOX element"
/rpt_type=dispersed
/rpt_unit=Box_ABBBC
353. .598
/gene="spr0485"
353. .598
/gene="spr0485"
/codon_start=1
/transl_table=11
/label-spr0485
/product="Hypothetical protein"
/protein_id="AAK9289.1"
/db_xref="GI:15458056"
/translation="MEVMDNITIDVSIPIVAEYVDKHPVEILIVELGFRPLNPLMRN
TVGRKVSIAKQGSKLACTPMDKIYRTILEANGYEYIGLD"
598. .1932
/gene="spr0486"
598. .1932
/gene="spr0486"
/codon_start=1
/transl_table=11
/label-spr0486
/product="Conserved hypothetical protein"
/protein_id="AAK9290.1"
/db_xref="GI:15458057"
/translation="MTDERHILRLDILIELHNGASPEVSODPRDAPFTGSAIEISIM
EHLNMSDQVTEPDYEMELCDVIANLFGNAIKGEVLDTEHPRPYRKRKEENLAIRA
ALIRIRLDITIELMEDEMLAKRGVLKQMGVGFQFDIHQKKELEFPIMERGH
DSPPKVMWGVDQIRLEFYALTATSLPEVSISSVKEAFEAFAFEESIKFEEESIL
LMLLESFTODDWIQIAESDAYAVALIRPESEKVFBERQSFIEKIAEBYVADTAEGL
QVOQVIDTEGHFTITFTPEKEAVLDRSQAFNGVSEQANLNLHNPWEITFV
NKEDIRYOYNDNPADENMFKRTPSGVGRNVELCHPKYLDKAKATMKGIRECTNKKY
EMWFKSESRGKPHITTYAAVHDEGEFGVLEYQDIQPREIDTYFQGLE"
1942. .2196
/gene="spr0487"
1942. .2196
/gene="spr0487"
/codon_start=1
/transl_table=11
/label-spr0487
/product="Hypothetical protein"
/protein_id="AAK9291.1"
/db_xref="GI:15458058"
/translation="MSYEOEFMKKEFANVNTQIMDMAHKSQKYVEDQDERAKDA
MIRYESRLDAYOFLKGFENFKAGKGFHDLPEGLFERNT"
2292. .2687
/gene="spr0488"
2292. .2687
/gene="spr0488"
/codon_start=1
/transl_table=11
/label-spr0488
/product="Hypothetical protein"
/protein_id="AAK9292.1"
/db_xref="GI:15458059"
/translation="MSKILNRKKQLNGLRSGAESSYTYKVVDETCKKYYKRAEGSAS
AAKRAVSKRVEQAVEATKREQQAVANSVEDFANLGLPLDRAKTEYDEGIRKSASDFK

gene
CDS
3099. .3548
/gene="spr0489"
3099. .3548
/gene="spr0489"
/codon_start=1
/transl_table=11
/label-spr0489
/product="Hypothetical protein"
/protein_id="AAK9293.1"
/db_xref="GI:15458060"
/translation="MIGVARENAQAQIQYQKFTYINISDETSMLEAGACFISHQK
LERLVGHEIISERTQIPIIDACPVLDCVDRIVEDDGIQHFPAKLTLEPLVAPELLDE
KGFKNQLPAPFTYFMGDGYQRYRYRLDKRVDKGSFKARRKDKRN"
3535. .4092
/gene="spr0490"
3535. .4092
/gene="spr0490"
/note="Probable acetyl transferase"
/codon_start=1
/transl_table=11
/label-spr0490
/product="Conserved hypothetical protein"
/protein_id="AAK9294.1"
/db_xref="GI:15458061"
/translation="MARTLPDKIETTERLVLRVRYAADIIDYASLPEAVYPAGFP
PKTLEDEIYLYEHLIPERNQKENTLPAGYIVKGTDKTVGSVDENHREDDVLEIGY
TRHPDYWGGRGVYVEAARALIDLAFKDLGHIKTELTCFGYNLQSKRAERKLGFTLEARI
RDRKDYQNRCDLSLYGLKSEMEP"
4094. .5056
/gene="spr0491"
4094. .5056
/gene="spr0491"
/codon_start=1
/transl_table=11
/label-spr0491
/product="Hypothetical protein"
/protein_id="AAK9295.1"
/db_xref="GI:15458062"
/translation="MSDVKEEVSSISEKQROIDVEYAEIANDSDIETRLAYLEINNE
KRIVSIDIEPTKEIMSDQIFEIFIOKFNKMFPLFSDVSDFLSIKNKSEKLE
IREADVNRMHILISGKLFVDFENQIQOKSKSEEDDQIHGAFSGVGYDINPYRF
CHSLRNSQHTDPIPIEYKAVSFDEDTYVIDYITDIDYILNSFNKKKIKGELIKIQ
ETSKIDALVREYFNALETGNNKLELKNHNLVDIKSKLESLEKLKLSHRYIYSK
ISKYDILVNGNTWMSPLAAFAIEIEIYIELSKIGLVKIVKNSN"
5068. .7719
/gene="valS"
/note="spr0492"
5068. .7719
/gene="valS"
/ec_number="6.1.1.9"
/codon_start=1
/transl_table=11
/label-spr0492
/product="Valyl-tRNA synthetase"
/protein_id="AAK9296.1"
/db_xref="GI:15458063"
/translation="MSKESSKYNPAAVEPAGRYOKMLDADVFPKSGDOKAKPYSYIIP
PNVNGKLTGHAMDTTLODIIIRKRMGQFPTLMPCMDHAGIATQAKVEERLGEES
ISRYDGRSEPLTKVWEKDEYATTIKEQMGKGLSVDSRERFTLDESLSAVRKYV
VDLYRKGMVLYRGEFIIINWDPARTALSDIEVILHKDVGAFYHMMNMLEGSSALEAVT
TRPETFEGVVALVNPEDPRYKDLIGKNVILPIANKLIPVGDENHADPEFGVYVAKIT
PAHPMDPLVGGORHNLPOVINMDQGTNMLEVFEESGDRFAPARAAVVAKLBEIGALV
KTEKRVHSGHSERGVYVEPRLSTQWFKMQOLAKNANATANDOTEDKVFEPYPREMDT
FLQAMENVAIDWITSRLMMGHQIIPANYNADGEMVYGEBAPEBDQMTQODEVDYTPRSS
ALMPFSTMGWPEVDSDEFKRYFPSTLVTVGYDITFFVSRMTIQSLDEFGRQPOVNL
IHGLIRDEQGRKMSKSLNGIDPMQVITEKYGADALRWFLSNGSACQDQVRSYEEKDA
SWNFINKIWNISRYILMNGGLTLDVADHNVTYKAVLFGAGVNTDVMILHNLNETTAKV
TENPDEEFGVAGHILYNTIMEEFANMYVELTKEVLYSDNDCKDYITRSVLLYTTLDKI
LTLHSHIMEVTEIEJEGOYAREGSITYAAAPVNPAPEDOLAARTQGVESIKDLIRAVNA
RAEVNVAAPSKPTIILVKTSDSLEAFPNNSNVAIKRTFPEHLELASTIPAPELAMSS
VTGGAETFLPLVDLLNVEEELARLEKELEKADKWKQELDMWGKRLSNRFANARPEYVQK
EKDKADYQAKYDVYVARIDEKKIKLV"

```

gene      7774. .8544
          /gene="spr0493"
CDS       7774. .8544
          /gene="spr0493"
          /codon_start=1
          /transl_table=11
          /label="spr0493"
          /product="Conserved hypothetical protein"
          /protein_id="AAK9297.1"
          /db_xref="GI:15458064"
          /translation="NMKVMKSHPHYELNLIGYGLAKFKDKLFKEFCOSKSEFYR
          VVSLGIAETGVKNNRMDLEPYFDNNRKGMMOKAEVYERKDLIDMFGENEHSYA
          EIVKMLASBKGKGTITVEKPIVTRFKRIODETGMKAEANFYLIHFDEKREKGGOLT
          DARLVGDYDROVDNOEYSLAEVKIRKSKGVRTATAEFEVKEKFDLISLVTN
          LDDIPKLVILIDNLFERKNIINKETIEYSVEDLY"
          8827. .9630
          /gene="spr0494"
          8827. .9630
          /gene="spr0494"
          /codon_start=1
          /transl_table=11
          /label="spr0494"
          /product="Hypothetical protein"
          /protein_id="AAK9298.1"
          /db_xref="GI:15458065"
          /translation="MOPYNIIDNPNTSYAKRDMRIGRGLOKVDNLVSAYESLAE
          KOSRGLTYEUYEDATYAHHTIDASTERADLVSLKIVELSHRFSEFATLILAKH

Alignment Scores:
Pred. No.:      11.9      Length:      10292
Score:          133.00     Matches:      106
Percent Similarity: 33.20%   Conservative: 64
Best Local Similarity: 20.70%  Mismatches: 164
Query Match:      3.58%     Indels:      178
DB:              1         Gaps:      25

US-10-008-355-2 (1-712) x AE008428 (1-10292)

QY 192 AspyAlphe-----LysAspyAlaArgMetValPheAlaPro 203
    |||||
DB 5143 GATGTTTTCAGCCTTGAGCGCATCAAAAGCCTTAATTCACGTTATTCACCA 5202
    |||
QY 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr 223
    |||
DB 5203 CCAACCTTACAGGTAACTT-----CACCTT 5229
    |||
QY 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
    |||
DB 5230 GGT-----CACGCT 5238
    |||
QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
    |||
DB 5239 TGGGATTCAC--ACTTTGACAGATATATATCATCCGTCAAAACGACGACAAAGGTTTGAT 5295
    |||
QY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
    |||
DB 5296 -----ACCCCTTGCCCTCCTGGATGGATGACCAACGACGAGGATGGCCACT 5337
    |||
QY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
    |||
DB 5338 CAGGCTAAGGTTGAGAGCCTTGCGTGTGAGGCAATTCCCGCTATGACCTTGTCGT 5397
    |||
QY 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
    |||
DB 5398 GAGTCTTTTTCAGCAAGAGCTGGGAATGGAAAGAC-----GATATATGCCACT 5445
    |||
QY 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
    |||
DB 5446 ACCATTCAGAGCAACATGGGCAAGATGGGCTCTCTGTAAGCTTTCTCGGACCGTTTC 5505
    |||
QY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluAlaArg 357
    |||
DB 5506 ACTCTTGACGCAAGGTTTGTC-----AAAGCTGTTGCTAAG 5541

```

```

QY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
    |||
DB 5542 GTCTTTGTGACCTTACCAAGAAAGGCTGGATCTACCGTGATGAGTTATATCACTACG 5601
    |||
QY 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
    |||
DB 5602 GACCCAGACACTGGCAGACGCCCTTTCTGATATATGAGGTATTCACAAAGATGTGGAAAGT 5661
    |||
QY 385 -----GluGlyAlaLysAlaAsnArgGluMetThr 394
    |||
DB 5662 GCCTTACACACATGATATACATGCTGTGAGAGATGGTTCACGGCCCTTGAAGTTGTAC 5721
    |||
QY 395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValAlaArgPheAlaGlnPheAlaAsn 414
    |||
DB 5722 ACTGCTCTGACATGATATGTTGGGAGCTT----- 5751
    |||
QY 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
    |||
DB 5752 GCGATTGCGGTTATATCCAGAA-----GACCGCGCTAC 5784
    |||
QY 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
    |||
DB 5785 AAGCATCTGATGTGTAAACAGTCATCCTTCCA---ATCGCTAATAAATCATCCCAATC 5841
    |||
QY 449 MetLeuAsp-----IleValArgArg 455
    |||
DB 5842 GTTGGAGATGACGACGAGATCTGAGTTGTGATCGTGTGTAAGAAACACACCTGCC 5901
    |||
QY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
    |||
DB 5902 CAGCATCCAAATGACTCTTGTGGTGGCCAAACGTCATACCTTGCACAAAGTCACATATG 5961
    |||
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
    |||
DB 5962 AACGACGACGAACTATGACAGCAGCTTCTTGAATTTTCAGGCGATG-----GAT 6012
    |||
QY 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
    |||
DB 6013 CGTTTGAACCTCGT-----AAGCAGCTCGTTGCTAAGTTGGAAGAAATC 6057
    |||
QY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
    |||
DB 6058 GGTGCCCTGTGCAAAATGCAAAAACGTGTCACAGTGTGTGCTACTCAGAACGACGACGT 6117
    |||
QY 532 -----IleGlnAla 534
    |||
DB 6118 GTGCTAGTTGAGCCACGCTTGTCTACTCAATGTTGCTCAAGATGACCAATTGGCTAAG 6177
    |||
QY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
    |||
DB 6178 AACGCCATTCGCCAACCAAGACACAGAGGACAAAG----- 6210
    |||
QY 555 ArgGluMetTyrProGlyAlaValAlaLeuProSer----- 565
    |||
DB 6211 GTGCAATTTTCACACCTCGTTTTCACAGTATACCTTCTTCATGATGAGAAATGTCCAC 6270
    |||
QY 566 AspAlaAsnPheThrMetAlaArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
    |||
DB 6271 GACGTGGTTATTCCTCTGCTGACGCTCTGCTGGGCTCAC-----CAATC 6312
    |||
QY 586 GlyAlaTrpTyrArgLysTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
    |||
DB 6313 CTGCTGTTGTRCAAT-----GCTGATGTGTAAATGTATGTGCGCGAAGAACTCCAGAA 6366
    |||
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
    |||
DB 6367 GGTGACGAGATGGAATCAAGCAAGAAAGCTCTTGAT 6402
    |||
RESULT 8
SPNEU1923/c      SPNEU1923      11362 bp      DNA      linear      HTG 11-JUL-2001
LOCUS           Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***
DEFINITION

```

in ordered pieces.

ACCESSION AL449945
 VERSION AL449945.1 GI:11545170
 KEYWORDS HTGS_PHASE2.
 SOURCE Streptococcus pneumoniae.
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.

REFERENCE 1 (bases 1 to 11362)
 AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldera,F., Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de Francesco,M., Pollssi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M. and Garcia-Bustos,J.F.
 TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate
 JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
 MEDLINE 21353329
 REFERENCE 2 (bases 1 to 11362)
 AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldera,F., Pollssi,A., Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de Francesco,M., Buell,G., Feger,G., Garcia,E., Peltsch,M. and Garcia-Bustos,J.F.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A., Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 COMMENT * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..11362
 /organism="Streptococcus pneumoniae"
 /serotype="19F"
 /db_xref="taxon:1313"
 /clone="G54"

BASE COUNT 3304 a 2518 c 2123 g 3414 t 3 others
 ORIGIN

Alignment Scores:
 Pred. No.: 13.4 Length: 11362
 Score: 133.00 Matches: 116
 Percent Similarity: 33.62% Conservative: 77
 Best Local Similarity: 20.21% Mismatches: 194
 Query Match: 3.58% Indels: 187
 DB: Gaps: 26

US-10-008-355-2 (1-712) x SPNEU1923 (1-11362)

QY 131 lysllevalysvalthrsplyvalgluglylnleuylsglylthrspglumet 150
 DB 5039 AAAATAAATAATACGAATTGCTCGAGAAAGACGTGCTAAATTCMAAATAAATA 4980
 QY 151 giutrgleu-arglysalaglnlgluvalcysgnglnleualatylslygluasalaas 170
 DB 4979 AACAAATATGAGAAAGACACACATGCTCTAAAGACTTTCATCTAAATACAAATCCAC 4920
 QY 170 peluasnglnleucysllevalgluprophetyrserasnaanglutyrphelulleva 190
 DB 4919 CGATGTTGAGGCGTGCCTTACCAAAATGCG-----CTTCA 4884
 QY 190 llyraspyalphe-----lysaspvalargmetvalpheal 202
 DB 4883 TGCAGTGTTCACAGCCCTTCAGCGCATCAAAAGCGTAACTTATTCATGCTGATTCC 4824
 QY 202 aroproserservalglylyspheclglylaspthrspasntipmetlirproarghi 222
 DB 4823 ACCACCAACGTPACTGGGAACCTT-----CA 4797
 QY 222 sthrclgiaspbeservalpheargvaltyralaglyalaaspaanargproalagluty 242
 DB 4796 CCTTGT-----CA 4788
 QY 242 rserlyspaspaanlyspiofrylspyrovaltyrphelaalaalasermetcnglnty 262

DB 4787 CGCTGGGATFACG---ACTTTCAGAGATATCATTCGCTCAAAAACCGCATCGAGGCTT 4731
 QY 262 rlysalaspaepyrralametthrleclglypheprogliserthrspargtylleu-- 281
 DB 4730 TCAAT-----ACCTTGGCTTCAGGTATGACACACGCTGATTCG 4689
 QY 282 -thrsertrpglyvalgluaspargilaeluasnglnuasnnaenproargilegluvalar 301
 DB 4688 GACTACAGCGTACGTTAGAGACCGCTTACGTGTGAGGGATTATACCGTTAGACCTGAG 4629
 QY 301 gclly-----lleysgngllylleftrpysgluallametseralaspclal 317
 DB 4628 TCGTAGCTTCTTTCGACCAAGCTGTGGCAATGGAAGAC-----GAATATGC 4581
 QY 317 athrarilleystryrala--serlystryralaenlseralaaenlytrplysasnse 336
 DB 4580 CACTACACATCAAGAAACATAGGGCAAGATGGGGCTTCTGTAGACTATTCGCGAGCG 4521
 QY 336 rtleclmetasnaraglyleualargleuaspvalileglyarqlysarqalagluc 356
 DB 4520 TTTCACTCTTACGAAAGCTTTGTCA-----AAAGCTGTTCG 4485
 QY 356 uargalaphelaasptripilearqlys-----aanglylsservalvalty 372
 DB 4484 TTAGGCTCTTTGTGACCTTTACAAAGAACGCTGATCATCTGCGTAGGTTTATCATCAA 4425
 QY 372 rcllyasp-----Valleuserserleuglnulysalatyryls----- 384
 DB 4424 CTGGGACCCAGACGCTCGACAGCCCTTCTCATATTGAGGATTCACAAAGATGTAGA 4365
 QY 385 -----gluglyalalyalaasnargluwe 393
 DB 4364 AGCTGCTTCTACACATGATTAATACATGCTGAGATGATGCTTCCGCCCTTGAAGTTGC 4305
 QY 393 tthrtyrleuserglutthrleupheclglythrclvalvalarqphelaaglphela 413
 DB 4304 TACAACTGCTCGTAGACTATGTTGGGACGTT----- 4271
 QY 413 aasnalaaleualartraspnaasproaspalaahislaaglylleuylsserleuaspsly 433
 DB 4270 -----GCGGTGCGATTAATCCAGAA-----GACCCGCG 4242
 QY 433 strlylasp-----Tyrlauproserleuasparqlysalalleupr 447
 DB 4241 CTACAGGACTTGATGTGTAATAACGTATCCTTCCA---ATCGCTAATAAACCTATCCC 4185
 QY 447 calamelleuasr-----llevalar 454
 DB 4184 ATTCGTTGGAGATGAGACGACGATCTGAGTTGCTGCTGTAATAATCACACC 4125
 QY 454 garqarglleproulaasplyleuproaspillepheyasnaValilleasplylsp 474
 DB 4124 TGCCCAAGCTCAAAATGACTTCTTGTTGGCCCAAGCTATACTTGCCCAAGTCAACGT 4065
 QY 474 elysglaspthrlylstrylasphevalpheasplysservalvalprotyrse 494
 DB 4064 CATGACGACGACGGAACATGATGAGCTTCCTTGAATTTCAAGCATG----- 4013
 QY 494 rasplysphehisalametleuylssermetaspysglulyspheaialyalailegl 514
 DB 4012 -GATGCTTTGGAAGCTCGT-----AAGCAGCTGTTGCTTAAGTTGGAAGA 3969
 QY 514 ulysasproulaalavalgluleuserlysservalillealaalaalargala----- 531
 DB 3968 AATCGGTGCTTGTCAAAATGAAAACGTGTTCATTCAAGTGTCTCACTCAAGACGAC 3909
 QY 532 -----ilegl 533
 DB 3908 AGGTGCTGATGTAGCCACGCTTGTCTACTCAATGTTGCTGCACAGATGACCAATTTGC 3849
 QY 533 nalasplamelalaaenalatyrallleclulyscllylsarqleuphephelaagl 553
 DB 533 ----- 553

Db	3848	TAAGAGCGCCATTCCAAACGACAGACAGGACGACAG-----	3812
Oy	553	yleuArlgUmetLytrProglYarGalaleuProSer-----	565
Db	3811	----GTCGAATTTCACCCACCGCTTTCACAGATACCTCCCTCAATGATGAGAAATGT	3756
Oy	566	----AspAlasAnphenrhetArmetSerTYrGlySerIleLysGlyTYrGluProG1	584
Db	3755	CCAGACTGCGGGTTTCCTCTGTCAGCTTGTGGGGGAC-----CA	3714
Oy	564	naspGlyAlaTrpTYrAsnTYrHisThrGlyLysGlyValLeuGluLysGlnAspR	604
Db	3713	AATCCCTGCCCTGGTACAT-----GCTGATGCTGAATATGATGCGAGAGAAAGCTCC	3660
Oy	604	olysSerAspGluheAlaValGlnGluAsnIleLeuAsp	617
Db	3659	AGAAAGTGACGATGACTACGACGACGACGACGCTTGAT	3620
RESULT 9			
AE009858		14169 bp	DNA
LOCUS			Linear
DEFINITION			BCT 16-JAN-2002
ACCESSION			AE009858
VERSION			AE009858.1
KEYWORDS			GI:18160615
SOURCE			Pyrobaculum aerophilum
ORGANISM			Pyrobaculum aerophilum
REFERENCE			1 (bases 1 to 14169)
AUTHORS			Fitz-Gibbon,S.T., Laderer,H., Kim,U.J., Stetter,K.O., Simon,M.I. and Miller,J.H.
TITLE			Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum
JOURNAL			Proc. Natl. Acad. Sci. U.S.A. 99 (2), 984-989 (2002)
PUBMED			11792869
REFERENCE			2 (bases 1 to 14169)
AUTHORS			Fitz-Gibbon,S.T., Laderer,H., Kim,U.-J., Stetter,K.O., Simon,M.I. and Miller,J.H.
TITLE			Direct Submission
JOURNAL			Submitted (12-DEC-2001) Microbiology and Molecular Genetics, University of California, Los Angeles, 405 Hilgard Ave, Los Angeles, CA 90095-1489, USA
FEATURES			Location/Qualifiers
SOURCE			1..14169
gene			/organism="Pyrobaculum aerophilum"
CDS			/strain="IM2"
gene			/db_xref="taxon:13773"
CDS			144..1019
gene			/gene="PAE2116"
CDS			144..1019
gene			/note="Protein fate; Protein modification and repair"
CDS			/codon_start=1
gene			/transl_table=11
CDS			/product="methionine aminopeptidase"
gene			/protein_id="AAL63956.1"
CDS			/db_xref="GI:18160616"
gene			/translation="MLRLQVGDVVNKKALYALDLPDMPLELCVRAEAFIRAND AKPAFPVNVSVNAHYHTAKRGDSIKTPKGLIVIDVAGQDGYIVRAATVYLGPV FNNLQKAAKSALESAINAKPGIKAWQIGEVVERVYKSGFLSPIYNLGHKVERILH AGVITNYSKATASQALAPGDVYATEPFAATNGEYVVGSGEITIRLARMHKNYOH. IDILISAEGRLPPTPMPQPLDDAIRALKAGVYLGVEVLYVERSKGVAQPEDTIYV GENEVPIYKTELL"
CDS			1060..2709
gene			/gene="PAE2117"
CDS			1060..2709
gene			/note="Protein fate; Protein folding and stabilization"
CDS			/codon_start=1
gene			/transl_table=11
CDS			/product="Thermosome (chaperonin) alpha subunit"


```

Db 9999 TACATAACCGTATACGTGACGTGACGT-----AGCTACTAGCCCAATTCGAC 10046
Qy 348 ValIleGlyArgLysArgAlaGluGluAlaPhe----- 359
Db 10047 GGC---TTCGAGAAAAAGCCGACAGTAACTATCATTCAGCGCTTTGGCGGCAAG 10103
Qy 360 AlaSPTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeu 379
Db 10104 GCGGAGTCAACCAAGATGCGCAATGACCTGCTTAC----- 10142
Qy 380 GluLysAlaTyrLysGluGlyAlaLysAlaAsnArgLysLeuMetThrTyrLeuSer----- 397
Db 10143 -----CTCCACACGCGCGCATTAATTCGATTCGCCATGACAGCGGTAAATGCGTGTG 10196
Qy 398 -----GluThrLeuPhe-----GlyGlyThrGluValAlaArgPheAlaGlu 411
Db 10197 AAGCGTTTGTGTTACCAACTTACAGTAAAGGCGCTGATCAGCGAGATACATACGAACAG 10256
Qy 412 PheAlaAsnAlaLeuAlaThrAsnProAspAla-----HisAlaGlyIleLeuLysSer 429
Db 10257 TTGCTGAGGACATACGTACAGGCCCAACACCGCTCACTATGCGGGGTGAGTTCTCG 10316
Qy 430 LeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMet 449
Db 10317 GTTATCACAAAGACGAGAGTAATACGATTTTAAATAAGTATACCCAGGC--- 10373
Qy 450 LeuAspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnVal 469
Db 10374 -----ANCGAGCGCTCTCAAAATGCCC 10394
Qy 470 IleAspLysLysPheLysGly-----AspThrLysLys 480
Db 10395 GTTGTGACCGCCCTAAAGCAAGAGGCTGAAGAGCGTTACACTTCACTGTCGAAGAG 10454
Qy 481 TyrAlaAspPheValPheAspLysSerValAlaProTyrSerAspLysPheHisAlaMet 500
Db 10455 TACGGGACACTACGAATTCACGTGCGTGTGAATCGATACAAAGAGCTTGGAGAGCCTTG 10514
Qy 501 LeuLysSerMetAspLysGlyLysPheAlaLysAlaIleGluLysAspProAlaValGlu 520
Db 10515 GCACGATGCTGGCTGAGAAAGGAGAGTACTACGCCATTCAT----- 10556
Qy 521 LeuSerLysSerValIleAlaAlaAlaArgAlaIleGluAlaAspAlaMetAlaAsnAla 540
Db 10557 ---GACAGAAAGCGCTGAATACAGCGTCAAGCGGAGCAGCAAAAGCAGGTGTAATACC 10613
Qy 541 TyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgLysMetTyrProGly 560
Db 10614 -----CTAAAAACCGCTGGCGTGGAGGAG----- 10637
Qy 561 ArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGly 580
Db 10638 -----GATACACACTTACGCTGAGATGAG----- 10661
Qy 581 TyrGluProGlnAspGlyAlaTyrTyrAsnTyrHisThrThrGlyLysGlyValLeuGlu 600
Db 10662 -----GACGGA---TACTACGTAAATTCGTAATTAATGAAGCGCTCCGCGAA 10706
Qy 601 LysGln-----AspProLysSerAspLysPheAlaValGlnGluAsn 614
Db 10707 ATCCAGCGTATGGCGCTAAACGGCGAGCGTGAAGCGTGGAGCGCTTTATA-----AGGAG 10760
Qy 615 IleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyLysGlnLeuHis 634
Db 10761 CTTGAGGATGCTGTAAG---CCCGCGCACGCGCAAAACCAACGATGAAGGCGATTTGAA 10817
Qy 635 IleAlaPheLeuSerAsnAsnAspIleThrGlyLysSerGlySerPro-----Val 652
Db 10818 GTGTGACGCGCGGAGGAGGAG-----GGGAGCATTAATTTACCGCTAGAAAGTG 10868
Qy 653 PheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspLysAsnTrpGluAlaMetSer 672

```

```

Db 10869 CGCGAGAAAGGGCAACGAGTGAAGCCGCGGTGTGAGCTTGAAGTATGATTCGTAATG 10928
Qy 673 GlyAsp-----IleGlu 676
Db 10929 GCGGATCAACCCGTTGGCAGTGCAGCGGGGAGAGACTGCCCGCTGCGATTAATAGTGAG 10988
Qy 677 PheGluProAspLeuGlnArgThrIleSerValAspIleArgTyrValLeuPheMetIle 696
Db 10989 TACGACACAGAGGGGGAAGAGAG-----CACCTAAAGATAGTATGTTACTGCGTT 11039
Qy 697 AspLys 698
Db 11040 AAGAAA 11045

RESULT 10
A81135
LOCUS A81135 2652 bp DNA linear PART 21-JAN-2000
DEFINITION Sequence 1 from Patent EP0909818.
ACCESSION A81135
VERSION A81135.1 GI:6731609
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2652)
AUTHORS Brown, J.R. and Wang, M.
TITLE Valyl CRNA synthetase (EC 6.1.1.9) from Streptococcus pneumoniae
JOURNAL Patent: EP 0909818-A 1 21-APR-1999,
SMITHKLINE BEECHAM CORP (US)
FEATURES
source 1..2652
location/Qualifiers
BASE COUNT 733 a 612 c 621 g 686 t
ORIGIN
Alignment Scores:
Pred. No.: 3 68 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.818 Conservative: 62
Best Local Similarity: 20.70% Mismatches: 166
Query Match: 3.50% Indels: 178
Gaps: 25

US-10-008-355-2 (1-712) x A81135 (1-2652)
Qy 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
Db 76 GATGTTTTCAGGCTTCAGGCGATCAAAAGGCTAAGCCTTATTCATTCGTTATTCACCA 135
Qy 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProAlaGlnHisThr 223
Db 136 CCAAAAGTTACAGTAACTT-----CACCTT 162
Qy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
Db 163 GGT-----CACGCT 171
Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
Db 172 TGGGATACA---ACTTTCGAAGATATTATCATCCGTCAAAAACGATGCAAGTTTGTGAT 228
Qy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
Db 229 -----ACCCCTTGGCTTCCCTGGGATGAGCCACGACGAGGATTCGCACT 270
Qy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgTrpIleGluValArgGly 302
Db 271 CAGGCTAAGGTAGAGAGCGCTGCGTGTGAAGGCAATTTCCGCTATGACCTTGCTCGT 330
Qy 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
Db 331 GAGCTTTTCTTGACGAAGCTCGGAATGGAAGAC-----GAAATTCGCACT 378

```


OY	374	ASP-----ValleusserLeugluysalaatyrglucllyalays	388
Db	535	GACCAGCAGACTCGCACGCCCTTTCTGATGTAGTGATTCCACAGAAGTGAAGGT	594
OY	389	AlaasnArglumetThrTyreu-----	396
Db	595	GCCTTCACCATGATAATTACATGCTGGAAGATGGTTCACGGCTCCTGAATGCTACA	654
OY	397	-----SergluthrleupheglylthrcluvalValargphealaglnPhealaasn	414
Db	655	ACTGCTCCTGAGACCATCTTTGGGGACGTT-----	684
OY	415	AlaleualthrAsnProaspalaHislaagiyleuleysserLeuaspsylstyr	434
Db	685	GCGGTGGCGTCACACCGAA-----GACCCGGCGTAC	717
OY	435	LysAsp-----TyreuproserLeuasparLyssvalleuprofa	448
Db	718	AAGCACTGATTTGGTA AAAATGTATCCTTCCA---ATGCGTAATAATCACTCCCAATC	774
OY	449	Melleuasp-----lleValIargar	455
Db	775	GTTGGAGATGAGCACAGCATCTGATGTTGGTACTGTCGTGMAAATCACACCTGCC	834
OY	456	ArgileproAlasplyleuproaspIlephelysasnaIIleasplsyshelys	475
Db	835	CACGATCCAAAATGACTTCTTGTTGGCCACAGCTCATTAATCTGCCACAAATCAAGTCATG	894
OY	476	gIyAsprhrLysTyrrAlasphevalPheasplysservalValprotyrSerasp	495
Db	895	AACGACGACGAGAACATGATGACTTTGGCCTTTGGAATTTTCAGGATG-----GAC	945
OY	496	LysPheHIsAlamelleulysserMetasplysgluLysPheAllysAlalleulyls	515
Db	946	CGTTTTGAAGCTCGT-----AAGCAGTCGTGTCYAAGTTGGMAAGAAATC	990
OY	516	AspProAlaValglLeusserLyssevalIIleAlaIalaIargAla-----	531
Db	991	GGTGGCTTCGTCAAAATGAAAAACGTGTCCACAGTGTGGTCACTCAGAGCCTACAGGT	1050
OY	532	-----lleGlinaIa	534
Db	1051	GTTGGTGTGAACCTCGGTTGTCATCTCAATGCTTCGTCAGATGGACCAATTTGGCTAAG	1110
OY	535	AspalameAlaasnalatyralallecluysglylsArgleuphePhealagleyeu	554
Db	1111	AACGCAATTTGCCAACCAACACACACAGAGACAAG-----	1143
OY	555	ArglumetyrPrroglYArGalaleuprosr-----	565
Db	1144	GTCGAATTTACCCACCTCGCTTTCAACAGATACCTTCCTCAATGATGGAATAATGTCAC	1203
OY	566	AspAlaasnpherhrMetarMetserTyrglyserIllelysglytyrclupProclnasp	585
Db	1204	GACTCGGTATTCCTCGACAGCTCGGGGCTAC-----CAAATC	1245
OY	586	gIyalatrpTYrarstyrrHisThrThrelgylsglyValleugluLysclnasPrrlys	605
Db	1246	CCTGCTGTGTACAAAT-----GCTGATGGTGAAGATGTGTCGGGAGAAAGCTCCAGAA	1299
OY	606	SeraSpgIupheAlaValaglInclunsnlleuasp	617
Db	1300	GGTGACGATGTGACTCAGAGACGACGATCTTGAT	1335
RESULT 12			
E26074	LOCUS		
DEFINITION	Novel vals.	2652 bp	DNA linear PAT 07-FEB-2001
ACCESSION	E26074		
VERSION	E26074.1 GI:13025049		
KEYWORDS	JF 1999113577-A/1.		
SOURCE unidentified.			

Alignment Scores:

Pred. No.:	7-47	Length:	4141
Score:	129.00	Matches:	153
Percent Similarity:	32.15%	Conservative:	110
Best Local Similarity:	18.70%	Mismatches:	288
Query Match:	3.47%	Indels:	268
DB:	1	Gaps:	38

US-10-008-355-2 (1-712) x CLOCOLA (1-4141)

```

QY      2 GlnMetLysLeuLysSerIleLeuLeuGlyAlaAlaLeuLeuGlyAlaSerGlyVal 21
DB      1518 CAGATGCAAGGAAATATCTCCAAAACATATACCTTTGTGATGATGAATTTT-GTA 1576
QY      22 AlaLysAla---AspLysGly-----MetTrpLeuLeuAsn 32
DB      1577 GTAAAAGCTGGTGTAAAGTAAACGAGAAAGATAAAGATTATATTTGGCGCTCAAG 1636
QY      33 GluLeuAsnGlnGluAsnLeu-----AspArgMetLysGluLeuGlyPhe 47
DB      1637 GAAGTTAAAGCTCAATTCATGAGAGTACCTCAAAATGATAGGCTTTAGAGAGGAAAT 1696
QY      48 ThrLeuProLeuAspSerLeuTyrSerPheAspLysProSerIleAlaAsnAlaVal 67
DB      1697 CCAGATGATATTTTAACTGTTGTTATTAATCACTCACGAGAGATATAAGTTAATCGT 1756
QY      68 IlePheGlyGlyGlyCysThr-----GlyIleThrValSerAspGlnGlyLeuIle 84
DB      1757 ATAAATTAAGTAAATAGTATGATGATTAATGCTGATGATTAATGAAACATAGCACTTTC 1816
QY      85 PheThrAsnHisCysGlyTyrGlyAlaIleGlnSer---GlnSerThrValAspHis 103
DB      1817 TTTACTTATGAAGAAGACACACAGAGAAAGTATATATACATTAGAGAAATATATCCGTCT 1876
QY      104 Asp-----TyrLeuArgAspGlyPheValSerArgThrMetGlyGluLeuPro 120
DB      1877 GAATTACTACATCTCTCAAGGTAGATATGTA----- 1909
QY      121 IleProGlyLeu-----SerValLysTyrLeuArgLysIleValLysValIleThrAsp 137
DB      1910 GTTCCCTGGAATGGGGGAGAGAGAAATTCATCAAGAGAGGATTTCCTGATGTA 1969
QY      138 LysValGlyGlyLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGln 157
DB      1970 GAAGGAACAGCAGAGTCTTTGAGAGTTCACACTAGCACTGATGAAATTAACCAACAAATA 2029
QY      158 GluValCysGlnGluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuGlyIleVal 177
DB      2030 TCAGTTACACAGAGGTTAGCTTACGATAGAAATATAGAAATGCTTATATGATGTAATTA 2089
QY      178 GluProPheTyrSerAsnAsnGlyTyrPheLeuIleValIleTyrAspValPheLysAspVal 197
DB      2090 CATGCTAAATATGCTCAATGGATTTCTAT----- 2119
QY      198 ArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrp 217
DB      2120 -----AATTATGATTTCTCTATCAAACTTC 2146
QY      218 MetTrpProArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsn 237
DB      2147 ATGTACAAACAATTAACATGGGAATGTT-----AAAT 2176
QY      238 ArgProAlaGlyTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaVal 257
DB      2177 AGATACAAATTAACATTAAGAAATAT----- 2203
QY      258 SerMetGlnGlyTyrLys-----AlaAspAspTyrAlaMetThrIle 271
DB      2204 GATGCTATCGGTTATTAAGATTATATGATCATCATAGTAGTAGATGATTAATGAT 2263
QY      272 GlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGluAspArgIleGlu 291
DB      --- 291

```

```

DB      2264 AATATCAAGACTATATGATCTTTATTA-----AACATATTCAT 2305
QY      292 AsnGluAsnAspProArgIleGlu-----ValArgGlyIleLysGlnGlyIleTrp 308
DB      2306 AACTTAGATGTTCTTATCTTACATGATGATGATGTAATGA----- 2347
QY      309 LysGluAlaMetSerAlaAspGlnAlaThrArg---IleLysTyrAlaSerLysTyrAla 327
DB      2348 CATGAAGCTAAGATATTAATGAATAACTAATGACATATAAAGAACTTCAAAATTAATA 2407
QY      328 Gln---SerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeu 346
DB      2408 GATCTTTCTAGTAATAGTTGAAAGCTCAATCTTACTACTTACATGAGAGAAACA 2467
QY      347 AspValIleGlyArgLysArgAlaGluGluAlaArgAlaPheAlaAspTrpIleArgLysAsn 366
DB      2468 TATGTAGGGGGAAGAAAGTCAAGGGGAAGAAAT-----GACTGGAAGATATGAAT 2518
QY      367 GlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlyLys----- 381
DB      2519 TCTAAG-----TTAATGATATATGTTAAAGAAATTAATCAAAAGAGCTGAATGCT 2569
QY      382 -----AlaTyrLysGlnGlyAlaLysAlaAsnArgGluMetThrTyrLeu 396
DB      2570 TATAAACTGTTACCTGATACCTTTGTAACCATTAAGCATTAAGTAAAGTAAATGCTAATGCTT 2629
QY      397 SerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeu 416
DB      2630 TATGATGTTGTATTCATGGAATGATACA----- 2659
QY      417 AlaThrAsnProAspAlaHisAla-----GlyIleLeuLysSerLeuAsp 431
DB      2660 GATACAAATACGATGTTTATGTTAATAAGACCTTAAGGCTGTTTAAATCT----- 2713
QY      432 AspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAsp 451
DB      2714 -----GATTTCTTCACTA----- 2725
QY      452 IleValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAsp 471
DB      2726 ATAGTTGAAGAAAGAAATTTAT-----TTTGATGGAACAGATCA 2764
QY      472 LysLysPheLysGlyAspThrLysTyr----- 481
DB      2765 AAGGATGAAGATGGTGAATTCAAAGCTTATGTAATGGGACTTTGGAGATGAGAAAAATCT 2824
QY      482 -----AlaAspPheValPheAspLysSerValValProTyrSerAspLysPhe 497
DB      2825 AATGAGCTAAAGCCACTCATAGTATATATAA-----ACTGGAGAAATAT 2869
QY      498 HisAlaMetLeuLysSerMetLys-----LysGluLysPheAla 510
DB      2870 GAAGTAAATTTAAACCTTACAGATATAAAGGTGAATTAATCTGAAGATAAAGATA 2929
QY      511 LysAlaIleGlyLysAspPro----- 517
DB      2930 AAAGTAGTGAAGATTAACCTGTTGAAGTTAATAATGAAGAGAGCAACCAATGATTTT 2989
QY      518 -----AlaValGluLeuSerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAsp 535
DB      2990 GAAAGGCTTAACCAATTAATCTAATCTAATAGTTAAGGTACTTTATACAGAAAGAG 3049
QY      536 AlaMetAlaAsnAlaLeuTyrAlaIleGlyLysGlyLysArg----- 548
DB      3050 GATTATTACATTAATATTAATTTTGTATGTAAGTAAAGGCAAGTTAAATCACTCTT 3109
QY      549 -----LeuPhePheAlaGlyLeuArgGluMet 557
DB      3110 AATAATTTAATTTCACTAGATTAATCTTGACACACTTTATTAAGAGGCAACCTAAACAT 3169
QY      558 Tyr-----ProGlyArgAlaLeuProSerAspAlaAsnPheThr 570
DB      3170 TATGTTTATATGCAACTGGAATGATGACAGAAATTAAGGGTGAAAGAACTTAGAG 3229

```

QY 571 MetargmetserTyrglySerlleuysglyTyrgluProInaspGlyAlaTrpTyrasn 590
 DB 3230 CCTGGAAGACTACTACTATGATCTATATGATATCATCAATCAGAGCT----- 3280
 QY 591 TyrHisThrThnGlyLysGlyValleuGluLysGlnAspProLysSerAspGluPheAla 610
 DB 3281 TACACAGTAATGTAAAGAAACCTTAAAGAACTTAAAGAAACAAAGCATGCT 3340
 QY 611 ValGlnGluAsnIleleuAspleuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsn 630
 DB 3341 ATAAAGAA-----GTTGAAAT 3358
 QY 631 GlyInleuHisIleAlaPheleuSerAsnAspIleThnGlyGlyAsnSerGlySer 650
 DB 3359 -----AACATGAT----- 3367
 QY 651 ProValPheAspLys-----AsnGlyAlaGluIleGly-----LeuAla 663
 DB 3368 -----TTTGATTAAGCTATGAAAGGTAGACAGTAAATGCAAAATGTGGAACATTAGC 3421
 QY 664 PheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuAlaArg 683
 DB 3422 AATGATGATCTTAAGGATTTATATGATATGATATCAAAATCCAAAGTACTTAACATA 3481
 QY 684 ThrIle-----SerValAspIleArgTyrValleuPheMetIleAsp 697
 DB 3482 GTAGTTGAACCTAGATTAATATAAATGAACCTGTTATTATATTCAGCGGAT 3535

RESULT 14
 BD003707/c 6004 bp DNA linear PAT 31-JAN-2002
 LOCUS BD003707
 DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
 ACCESSION BD003707
 VERSION BD003707.1 GI:18631668
 KEYWORDS JP 2001501833-A/27.
 SOURCE unclassified.
 ORGANISM unclassified.
 REFERENCE 1 (bases 1 to 6004)
 AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
 Fannon, M. and Dougherty, B.A.
 TITLE Polynucleotide of Streptococcus pneumoniae and sequence
 JOURNAL Patent: JP 2001501833-A 27 13-FEB-2001;
 HUMAN GENOME SCIENCES INC
 COMMENT OS Unidentified
 PN JP 2001501833-A/27
 PD 13-FEB-2001
 PF 30-OCT-1997 JP 1998520718
 PR 31-OCT-1996 US 60/029960
 PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
 STEVEN C BARASH,
 MICHAEL FANNON, BRIAN A DOUGHERTY
 PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
 PC C12N1/21,
 PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
 G06F15/40
 CC Strandedness: Double;
 CC Topology: linear;
 FH key
 FT source
 FT location/Qualifiers
 FT location/Qualifiers
 FT location/Qualifiers

FEATURES
 source 1..6004
 location/Qualifiers
 /db_xref="taxon:32644"
 BASE COUNT 1733 a 1296 c 1115 g 1860 t
 ORIGIN

Alignment Scores:
 Pred. No.: 11.8 Length: 6004
 Score: 129.00 Matches: 116
 Percent Similarity: 33.62% Conservative: 77

Best Local Similarity: 20.21% Mismatches: 194
 Query Match: 3.47% Indels: 187
 DB: 6 Gaps: 26
 US-10-008-355-2 (1-712) x BD003707 (1-6004)
 QY 131 LysIleValLysValThrAspLysValGluGlnLeuLysGlyIleThrAspGluMet 150
 DB 4315 AAAATTAATTAATAGCAATTTCTGCAAGAAAGACGCTGTAAAGAAATTCAAATTAATA 4256
 QY 151 GluArgLeu-ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGluAsnAlaAs 170
 DB 4255 AACAAATAGAGAAAGACACACACATGCTTAAAGAACTTCACTTAATCAATCCAGC 4196
 QY 170 pGluAsnGlnLeuLysIleValGluProPheTyrSerAsnGluLysPheLeuIleVal 190
 DB 4195 CGAGGTGAGGCTGCTGCTTACCAAAATG-----CTTGA 4160
 QY 190 LTrAspValPhe-----LysAspValArgMetValPheAl 202
 DB 4159 TCGGATGTTTTCAGGCTTCAGCGATCAAAAGGCTTAAGCTTATTCATCGATTC 4100
 QY 202 aProPheSerValGlyLysPheGlyLysAspThrAspAsnTrpMetTrpProArgH 222
 DB 4099 ACCACCAACCTAATCTGGAAACTT-----CA 4073
 QY 222 sThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyr 242
 DB 4072 CCTTGT-----CA 4064
 QY 242 rSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyr 262
 DB 4063 CGCTTGCGATACG--ACTTTGACAGATATCATCATCCGTCMAAAGCATGCAAGCGCTT 4007
 QY 262 rLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu-- 281
 DB 4006 TGAT-----ACCTTGTGCTTCCAGCTTGGACACACGCTGTGATTC 3965
 QY 282 -ThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValAr 301
 DB 3964 GACTCAGGCTAAGGTGAGGAGCGCTTACGTGGAGGCTTACGCTTATGACCTTAGG 3905
 QY 301 gGly-----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAl 317
 DB 3904 TCGTAGTCTTCTTGAAGAGGCTGGAATGGAAGAC-----GAATATGC 3857
 QY 317 aThrArgIleLysTyrAla-----SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSe 336
 DB 3856 CACTACCAATCAAGAAACAAATGAGGCAAGATGGGCTTCTGTAGACTATCTCGAGCG 3797
 QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 356
 DB 3796 TTTCACTCTTATAGAGTTTGTCA-----AAAGCTGTTCG 3761
 QY 356 uArgAlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyr 372
 DB 3760 TAAGGTCTTGTGAACCTTACAGAAAGGCTGATCATCTGAGTGTATATCAACAA 3701
 QY 372 rGlyAsp-----ValLeuSerLeuGlnLysAlaTyrLys----- 384
 DB 3700 CTGGAGCCACACGCTGCGACAGGCTTCTGTATAGATGAGTGTCAAGAGATAGA 3641
 QY 385 -----GluGlyValLysAlaAsnArgLume 393
 DB 3640 AGGTGCTTTCACACATGATTTACATGCTGGAAGATGTTACAGCGCCCTTGAAGTTGC 3561
 QY 393 tThrTyrLeuSerGluThrLeuPheGlyGlyTyrGluValValArgPheAlaGlnPheAl 413
 DB 3580 TACAACCTGCTCGAGACTATGTTGGGAGCTT----- 3547
 QY 413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspArgLys 433
 DB 3546 ---GCGGTTGCGGTATATCCAGAA-----GACCCGCG 3518

```

Qy 433 STYLYSASP-----TyrLeuProSerLeuAspArgLysValLeuPr 447
    :|||||
Db 3517 CTACAGAGACTTGATGGTAAAGCGTCATCTTCCA---ATCGTAATAACATCATCC 3461
Qy 447 OAlaMetLeuAsp-----IleValar 454
    :|||
Db 3460 AATGCTTGAGATGACGACGAGATCCTGAGCTTGACTGTCGTCGTAAGATCACC 3401
Qy 454 GARARGLLEPrOAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPhe 474
Db 3400 TGCCACAGCATCAATGACTCTTGTTGGCCAAAGTCATACCTGACCAAGTCACAGT 3341
Qy 474 elyGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValAlaProTyrSe 494
Db 3340 CATGACGACGACGAGCACTAGATGAGCTTGCTTAATTTTCAGGCATG----- 3289
Qy 494 rAspLysPheHisAlaMetLeuLysSerMetAspLysLysPheAlaLysAlaIleGl 514
Db 3288 -GATCGTTTGAAGCTGCT-----AAGCCAGTCGTTGCTAAGTTGGAGA 3245
Qy 514 ulysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
Db 3244 AATCGGTGCCCTTGCAAAATCGAAACGTGTCACAGTGTGTGCTCAGAACGTAC 3185
Qy 532 -----IleGl 533
Db 3184 AGTGTCCTAATTGAGCCAGCCCTGTCCTACTCAATGTTGCTCAAGATGACCAATTTGGC 3125
Qy 533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGl 553
Db 3124 TAAGAACCCATGCTGTACCAAGACACAGAGACAG----- 3088
Qy 553 yLeuArgLysMetLysProGlyArgAlaLeuProSer----- 565
Db 3087 ----GTCGAATTCTACCCAGCTGTTCAACGAAACCTTCCTCAATGATGAAATGT 3032
Qy 566 ----AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGl 584
Db 3031 CCAGCAGCGGGTTATCTCTGCTCAGCTTGTTGGTGGGTAC-----CA 2990
Qy 584 nAspLysAlaTPrPyrAsnTyrHisThrGlyLysGlyAlaLeuGluLysGlnAspPr 604
Db 2989 AATCCCTCGCTGTACAT-----GCTGATGTCGAATGATGATGCGGCAAGACGTC 2936
Qy 604 oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 2935 AGAAGGTGACGAGTGAAGCTCAGACGAGACGCTTGGAT 2896

RESULT 15
AE007366 11280 bp DNA linear BCT 31-AUG-2001
LOCUS Streptococcus pneumoniae TIGR4 section 49 of 194 of the complete
DEFINITION genome.
ACCESSION AE007366 AE005672
VERSION AE007366.1 GI:14972031
KEYWORDS
SOURCE
ORGANISM Streptococcus pneumoniae TIGR4.
          Streptococcus pneumoniae TIGR4.
          Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
          Streptococcus.

REFERENCE
  1 (bases 1 to 11280)
  Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
  Peterson, S., Heidelberg, J., Deboy, R.T., Haft, D.H., Dodson, R.J.,
  Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
  Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,
  Holtzapple, E., Khouri, H., Wolf, A.M., Uterback, T.R., Hansen, C.L.,
  McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T.,
  Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O.,
  Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and
  Fraser, C.M.
  Complete genome sequence of a virulent isolate of Streptococcus
  pneumoniae

```

```

JOURNAL Science. 293 (5529), 498-506 (2001)
MEDLINE 21357209
PUBMED 11463916
REFERENCE 2 (bases 1 to 11280)
AUTHORS Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
          Peterson, S., Heidelberg, J., Deboy, R.T., Haft, D.H., Dodson, R.J.,
          Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
          Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,
          Holtzapple, E., Khouri, H., Wolf, A.M., Uterback, T.R., Hansen, C.L.,
          McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E.,
          Loftus, B.J., Uwal, M.L., Yang, F., Smith, H.O., Venter, J.C.,
          Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
TITLE Direct Submission
JOURNLAB Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
FEATURES
source Medical Center Dr, Rockville, MD 20850, USA
          Location/Qualifiers
          1..11280
          /organism="Streptococcus pneumoniae TIGR4"
          /strain="TIGR4"
          /db_xref="taxon:170187"
          156..506
          /gene="SP0557"
          156..506
          /gene="SP0557"
          /note="identified by match to PFM protein family HMM
          PF02033"
          /codon_start=1
          /transl_table=1
          /product="ribosome-binding factor A"
          /protein_id="AAK74713.1"
          /db_xref="GI:14972032"
          /translation="MANHFRDVRGMEIKREVENIIOKKVRDPVGVGTTIDVOMLGD
          ISVAKVYVTTISNLSADNKAQIGLEAKGTITRELGRLNLIKVIPIPLTVKKESTIEY
          GNRIDEMRLNDKN"
          749..1564
          /gene="SP0558"
          749..1564
          /gene="SP0558"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=1
          /product="conserved hypothetical protein"
          /protein_id="AAK74714.1"
          /db_xref="GI:14972033"
          /translation="MIRCKREISIVYABODIAMOVLOQVVKLPYVVDKSLPYDKF
          SKLDPKIDPTLLEQGFPTLLSDETIDRVANACIRNVLAAGTSLVAGIPGLAMAI
          TTPADVAQFYAFSLKLAQELGYLYGEDIMASRELSDEADQNTLLTLVGLVGNCTA
          ALRVGSIITAKQYMKIVPNKALKITLPIILKLVKIFGVNLTGGLAGMGFTPI
          LGGIIISGITTFATKPMGESLQKELSKLVNVSEVOYQEDVETIRKEAIIKGE"
          1566..1748
          /gene="SP0559"
          1566..1748
          /gene="SP0559"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=1
          /product="hypothetical protein"
          /protein_id="AAK74715.1"
          /db_xref="GI:14972034"
          /translation="MNPICAFKIYGNVFLTVGVKVMKTIKRAHVVGIGKLEIAD
          KIMDTARMLIKPEERE"
          complement(1701..1934)
          /gene="SP0560"
          complement(1701..1934)
          /gene="SP0560"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=1
          /product="hypothetical protein"
          /protein_id="AAK74716.1"
          /db_xref="GI:14972035"
          /translation="MEFLVLCNDVHLDFKEPTIOYLKHYLVKQOLDCKRDPHPKEF
          HNPNNRDKRNSKTKKISFSLMLNPSRIH"

```

gene 2196. .2441
/gene="SP0561"
2196. .2441
/gene="SP0561"
/note="similar to GB:AE000782; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK74717.1"
/db_xref="GI:14972036"
/translation="MEVMDNIDVSIYVAEVDKHPVEIIVELGFKPLANPLMRN
TVGRNVLSKQSGSKLGTMPMDKIVRTLLEANGVEVIGLD"
2441. .3775
/gene="SP0562"
2441. .3775
/gene="SP0562"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK74718.1"
/db_xref="GI:14972037"
/translation="WTDRIHLRDLILLEHNGASPSVODRFDATGTGSAIEISLM
EHELNNSGTVFEDVMEICDVHANLFENALKGVEVSTEPHGHVVFKEENALARA
ALIRLRLLDIYESMEDEMELEMKRGVIRMGVGOEDIDYORKELFPIIMRYGH
DSPPVGVGVDOQIRELFQPTALLTKAKSLPEVSISSVKEAPAFATFESMIFKEESTL
LILLESFTQDDMLQIAESDAYGAIIRPEKVPPEQSTIEKIDEPYQDTAGC
QVQYQIDYDPGHFTITFTPKEREAVLDHSSQAFNGSLVSEANLILNHLPMTFIV
NKEEDFOYVNDNTPADEMIFKRTPSQVGRNVELCHPKPYLDKVTIKKGREGSKDY
EMWFSSESGKFEVHTTAAVHDEDEGEFQGVLEYQYDIQPYREIDTYFERGLE"
3785. .4039
/gene="SP0563"
3785. .4039
/gene="SP0563"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74719.1"
/db_xref="GI:14972038"
/translation="MSYDQEFMKFEFVAVNTQIMINDMAHKSQKXYEEDDERAKDA
MRYESRLDAYOFLGKPFENFVGKGFHDLPGLFERNY"
4135. .4530
/gene="SP0564"
4135. .4530
/gene="SP0564"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK74720.1"
/db_xref="GI:14972039"
/translation="MSKRLNRRKQRLRNGLRAGAFSSVTYRKVDETKRVKRAEOSAS
AAGKAVSRKVEADVATEKQEQAVANSVEDPAAAMIGLIDPRAKTFEDESICKSPDK
NATEKELALKGIGPATIKKLKENGICKK"
4943. .5392
/gene="SP0565"
4943. .5392
/gene="SP0565"
/note="similar to GP:8927331; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK74721.1"
/db_xref="GI:14972040"
/translation="MIGVARENAAEOIKQYKQFTVINISDETSMIAEQAQGISHOEK
LERLVAHEISRTQPTILDACPLVLCORNDRIYEEGICHIAPKILIERLVAPELDE
KGHFENQLEAFYTFMGDTQRYRVLDRKVDKMSFTKKARKDKGN"
5379. .5939
/gene="SP0566"
5379. .5939

[illegible]

Db 6734 TGAT-----ACCTTGGCTTCCAGGTATGACACCGCTGATTGC 6775
QY 282 -ThrserrgLyValGlnuaspArgIleGluuAnGluuAsnProArgIleGluValAr 301
Db 6776 GACTCAGGCTAAGGTCGCGCTTACGTCGTGAGGATTAACGGTATACCACTAGG 6835
QY 301 gGly-----IleGlnGlnGlyIleTyrPlySGluAlaMetSerAlaSpGlnAl 317
Db 6836 TCGTGAGTCTTCTTGACGAAAGTCGCGGAATGGAAGAAC-----GAATATGC 6883
QY 317 aThrArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTyrPlySAsnSe 336
Db 6884 CACTACCATCAAGCAATGAGGCGCAAGATGGGCTTCTGTAGACTATCTCTCGTAGCG 6943
QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 356
Db 6944 TTTCACTCTTGATGAAGTTGTCA-----AAAGCTGTCG 6979
QY 356 uArgAlaPheAlaAspTyrPleArgLys-----AsnGlyLysSerAlaValTyr 372
Db 6980 TAAGGCTCTTGACACTTACAAAGCAAGCTGATCTATCGTGTGATTCATCA 7039
QY 372 rGlyAsp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
Db 7040 CTGGGACCCAGCAGCTCGCACAGCCCTTCTGATATTGAGTGATTCACAAGCATGTAGA 7099
QY 385 -----GluGlyAlaLysAlaAsnArgLys 393
Db 7100 AGTGCCCTTACCACATGATTACATGCTGGAAGATGGTTCAAGCGCCTTGAAAGTTGC 7159
QY 393 tThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAl 413
Db 7160 TACACACTCGTCGAGACTGATGTTGGGAGCGT----- 7193
QY 413 aAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLys 433
Db 7194 ----GCGGTTGCGGTTAATCCAGAA-----GACCCGCG 7222
QY 433 sTyrLysAsp-----TyrLeuProSerLeuAspArgLysValLeuPro 447
Db 7223 CTACAGGACTTGATGTTGTAACCAACGTCATCTTCCA--ATCGCTAATAACTCATCC 7279
QY 447 oAlaMetLeuAsp-----IleValAr 454
Db 7280 AATCGTTGGAGATGAGACACCATCTGACCTTGACTGTCGTGAAATACACACC 7339
QY 454 gArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPh 474
Db 7340 TGCCACAGATCCAAATGACTTCTGTGGCCAAAGTCATACTTGCACACAGTCACAGCT 7399
QY 474 eLysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSe 494
Db 7400 CATGAACGACGAGCAATGAAATGAGCTTGGATTTTCAGGCATG----- 7451
QY 494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleG 514
Db 7452 -GATCTTTTGAAGCTGCT-----AAGCAGTCGCTCTTAAGTTGGAAGA 7495
QY 514 uLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
Db 7496 AATCGGTGCCCTTGTCAAAATCGAAACGTCACAGTGTGTCACACAGAACGTAC 7555
QY 532 -----IleG 533
Db 7556 AGGTGCTAGTGAAGCCAGCTTGTCTACTCAATGGTTCGTCACAGATGACCAATTTGC 7615
QY 533 nAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaG 553
Db 7616 TAAGAAAGCATTCCTAACCACAGACACAGAGACAG----- 7652
QY 553 yLeuArgGluMetTyrProGlyArgAlaLeuProSer----- 565
Db 7653 ----GTGAATTTGACCCACCTGTTCAACGATACCTTCTTCATGATGAAATGT 7708

QY 566 ----AspAlaAsnPheThrMetLArgMetSerTyrGlySerIleLysGlyTyrGluProG 584
Db 7709 CCAAGACTGGGTATCTCTGTCAGCTTGTGGGGTCAC-----CA 7750
QY 584 nAspGlyAlaTyrPyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspPr 604
Db 7751 AATCCCTGCCCTGGTACAT-----GCTGATGGTGAATGTATGTGCGCGAAGAGCTCC 7804
QY 604 oLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
Db 7805 AGAAGTGACGATGAGACTCAGAGCAAGAGAGCTTGGAT 7844

Search completed: October 17, 2002, 23:36:03
Job time : 2700 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:39:21 : Search time 217 Seconds

(without alignments)
5633.378 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 MOKKLSILGALLGASG.....LFMIDKMGCPRIQLKLI 712

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1736436 segs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlh
-O=/cgn2_1/USPRO.spool/US10008355/runat_16102002_121152_27685/app_query.fasta_1.903
-DB=N.Geneseq_032802 -QFMT=fastap -SUFFIX=ring -MINMATCH=0 -1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsu2 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355 -ECGN_1_1_65_etunal_16102002_121152_27685 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGESOURCY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq_032802: *
1: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	134.5	3.6	2853	20 AAX91754	Porphyromonas ging
2	134.5	3.6	2859	20 AAX91621	Porphyromonas ging
3	133	3.6	2652	23 AAS55647	Streptococcus pneu
4	130	3.5	2652	20 AAX37027	S. pneumoniae valis
5	130	3.5	2652	20 AAX32814	S. pneumoniae valis
6	129	3.5	2652	21 AAA14369	DNA encoding strep
7	129	3.5	6004	19 AA052160	Streptococcus pneu
8	127	3.4	2652	23 AAS55949	Streptococcus pneu
9	123.5	3.3	2142	20 AAS51948	M. catarrhalis str
10	119.5	3.2	3271	22 AA164839	XKN-100 coding seq
11	117.5	3.2	1664976	19 AA042063	Methanococcus jann
12	117	3.1	1830121	17 AA042063	Haemophilus influe
13	116.5	3.1	15424	21 AAC81763	Porcine lelystad v
14	116.5	3.1	15424	21 AAC81764	Porcine lelystad v
15	116	3.1	3598	21 AAC77776	Human cancer assoc
16	115	3.1	4853	17 AAT13139	Alpha-D-glucosyltr
17	114	3.1	2136	17 AAT11244	Neisseria meningit
18	114	3.1	3420	22 AAH14452	Human cDNA sequenc
19	114	3.1	3431	21 AAC76951	Human ORFX ORF2506
20	113.5	3.1	2856	23 AAS52731	E. coli DNA for ce
21	113	3.0	8930	19 AA022834	Haemophilus paraga
22	112.5	3.0	36471	21 AA081453	N. meningitidis pa
23	112.5	3.0	349980	21 AAF21611	Neisseria meningit
24	112.5	3.0	837096	21 AA081489	N. meningitidis pa
25	112	3.0	2175	23 AAS79571	DNA encoding novel
26	112	3.0	2175	23 AAS81718	DNA encoding novel
27	112	3.0	3285	22 AAH54398	S. epidermidis gen
28	112	3.0	56506	21 AA069168	Bacteriophage Dp-1
29	111	3.0	4411529	22 AA199682	Mycobacterium tube
30	110.5	3.0	3279	14 AA051556	Sequence encoding
31	110.5	3.0	12665	19 AA052267	Streptococcus pneu
32	110.5	3.0	15450	21 AA027809	North American por
33	110.5	3.0	349980	22 AAH68530	C. glutamicum codin
34	110.5	3.0	1230025	20 AAX91990	Nucleotide sequenc
35	110	3.0	2255	17 AAT11240	Neisseria meningit
36	110	3.0	2319	19 AA036920	Thermotoga maritim
37	110	3.0	2538	21 AA055217	C. symbiosum open
38	110	3.0	3279	14 AA050946	Sequence encoding
39	110	3.0	32998	21 AA055186	Cenarchaeum symbio
40	109.5	2.9	2082	21 AA050408	Streptococcus pneu
41	109.5	2.9	2319	18 AAT93691	Thermotoga maritim
42	109	2.9	1490	21 AA059611	Human secreted pro
43	109	2.9	6703	21 AA05436	Streptococcus pneu
44	109	2.9	15363	19 AA052272	Streptococcus pneu
45	109	2.9	349980	22 AAH41225	Pyrococcus abyssi

ALIGNMENTS

RESULT 1	
AAX91754	
ID AAX91754 standard; DNA; 2853 BP.	
AC AAX91754;	
XX	
DT 25-AUG-1999 (first entry)	
XX	
DE Porphyromonas gingivalis protein Pg67 encoding DNA.	
KX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;	
KW vaccine; antigenic; ds.	
XX	
OS Porphyromonas gingivalis.	
XX	
PN WO9929870-A1.	
XX	
PD 17-JUN-1999.	

XX 10-DEC-1998; 98WO-AU01023.
 XX 04-AUG-1998; 98AU-0005028.
 PR 10-DEC-1997; 97AU-0000839.
 PR 31-DEC-1997; 97AU-0001182.
 PR 30-JAN-1998; 98AU-0001546.
 PR 10-MAR-1998; 98AU-0002264.
 PR 09-APR-1998; 98AU-0002911.
 PR 23-APR-1998; 98AU-0003128.
 PR 05-MAY-1998; 98AU-0003338.
 PR 22-MAY-1998; 98AU-0003554.
 PR 29-JUL-1998; 98AU-0004917.
 XX (CSLC-) CSL LTD.
 XX Agius CT, Barr IG, Hocking DM, Margetis MB, Patterson MA;
 PI Ross BC, Rothel LT, Webb EA;
 XX WPI; 1999-385613/32.
 DR P-PSDB; AAY34536.
 XX Antigenic Porphyromonas gingivalis peptides for preventing
 PT gingivitis
 PS Claim 12; Page 244-245; 588bp; English.
 XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
 CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
 CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
 CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
 CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.
 XX
 XX Sequence 2853 BP; 736 A; 685 C; 691 G; 741 T; 0 other;
 SQ
 Alignment Scores:
 Pred. No.: 0.00723 Length: 2853
 Score: 134.50 Matches: 105
 Percent Similarity: 34.41% Conservative: 66
 Best Local Similarity: 21.13% Mismatches: 202
 Query Match: 3.62% Indels: 124
 DB: 20 Gaps: 23
 US-10-008-355-2 (1-712) x AAX91754 (1-2853)
 QY 221 ArgHisThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240
 DB 717 CGGTCACAGTCGCGTCAATCTGTATCGCCGATATCGCGTCCGATAGTCTTCTT 776
 QY 240 aglUtyrSerLysaspSnLysProTyrLysProValTyrPheAlaValSerMetG1 260
 DB 777 CGGCTATTTT-----CCCTTTCACCAAAATACTCTTCGCTATTTGATGCC 824
 QY 260 nglTYrLYsAlaAspAPtyrAlaMetThrLleGlyPheProGlySerThrAspArgTY 280
 DB 825 CAGGTACGAGAGGACAT-----CGCTATGATTT-----TA 857
 QY 280 rleUthSerTrrpLy-----ValGluAspArgLleGluAsnGluAsnAsnPr 296
 DB 858 TTTCAGGAAATGTCGATATATTTCCTTCAGCGACTATATCGAT----- 903
 QY 296 oArgLleGluValArgLy-----LleYsGlnGlyLleTrrpLysGluAlaMetSerAl 314
 DB 904 ----TTGGCATTCGTCGGGAGATCTTTCCAAAGGTCATTCAGCCCATC 959
 QY 314 aAspGlnAlaThrArgLleLysTyrAlaSerLysTyrAlaGlnSerAlaSnTrrpLy 334
 DB 960 GAAATATTAAGAGAGGTATATAGTACAAAGCGCTGCTC-----GAGCCAAATTA 1008

QY 334 sAsnSerLleGlyMetAsnArgLysLeuAlaArgLeuAspValLleGlyArgLysArgAl 354
 DB 1008 ----- 1008
 QY 354 aglUgluArgAlaPheAlaAspTrrpLleArgLysAsnGlyLysSerAlaValTrrpLyAs 374
 DB 1009 -----CTGTATCGAAGTCCGGGACAAATATCTGCGCGAGA 1046
 QY 374 rValLeuSer-----SerLeuGluLysAlaTyrLysGluLysAlaAs 390
 DB 1047 CTACAGCAAAACCCAGCTGTGAATATCCATGAGACACACAGTCAGATCCGAAGCCAA 1106
 QY 390 nArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValArgPheAl 410
 DB 1107 TCCT-----TTGCAAAAGTGTGCGCAATGTCAATTTTCCACCGGAG 1151
 QY 410 aglInPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla-----GlyLleLeuY 428
 DB 1152 CTATTTCCAGAAATTCGCTGAATACCCACCTATGATGTCAATGCGTACTGCTACGACACG 1211
 QY 428 sSerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAl 448
 DB 1212 AAGTTCGCGCGTACGATATTCGCGCAAGTTCCGCGTACTCTTTTCATTCAGGCTAG 1271
 QY 448 aMetLeuAspLle-----ValArgArgArgLleProAlaAspLysLeuProAspAl 465
 DB 1272 CATG---GATATACGCGAGAACATGCGCGATACGACGCGTACCTTCTGCGCAATCT 1328
 QY 465 ePheLysAsnValLleAspLys---LysPheLysGlyAspThrLysLysTyrAlaAspH 484
 DB 1329 TTGCATTAATATGTCACGCGGTATTCCTTTCACAGGAGACCCGCTAGACCGCGCGC 1388
 QY 484 eValPheAspLysSerValAlProTyrSerAspLysPheHisAlaMetLeuLysSerMe 504
 DB 1389 ATGGTACGAGAAGTGTAGTGGGTATTCGCGTATTCGCGTACCTCGCAATGTATCTTGACAA 1448
 QY 504 tAspLysGluLysPheAlaLysAlaLleGluLysAspProAlaValGluLeuSerLysSe 524
 DB 1449 AGAGAAAGATTTGCTCCAGACATCTCGCGCATTTGGAAGAAATGTGGCTCATTC 1508
 QY 524 rValLleAlaAlaAlaArgAlaLleGlnAlaAspAlaMetAlaAsnAlaTyrAlaLleG1 544
 DB 1509 CGTACCGATAGTTTGCATCTTCCCTTGTGATTAATATCTGACTATGGGGGTAA 1568
 QY 544 uLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyr-----ProGlyArg 561
 DB 1569 CTACAAATGAGTGGGTGATACGAAAGCATATCGGAAGTCTGGAATGAGATGAAGAAAC 1628
 QY 561 gAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerLleLysGlyTy 581
 DB 1629 ATTCTGCTTCGCGACGACGACCTATAAATTCGCGACACTGTAC-----GATTA 1676
 QY 581 rGluProGlnAspGlyAlaTrrpTyrAsnTyrHisThrGlyLysGlyValLeuGluLyu 601
 DB 1677 CAGTCTGTCGCGAGGC-----TTATGTACCACTTGTACGATATGTTCAAGCC 1724
 QY 601 sGlnAspProLysSerAspLlePheAlaValGlnGluAsnLleLeuAspLeu----- 618
 DB 1725 TTGGAACCT-----TTTTCCTTCGAGGGAATCTCATATGATCCGTCATCG 1772
 QY 619 -----PheArgThrLysAsnTrrpLyArg 636
 DB 1773 CTTCAGCCGCACTGTAGTTTCTCTATATGCGCGACTGACAAACGCGATATGTCCTT 1832
 QY 626 gTyrAla-----GluAsnGlyGlnLeuHisLleAlaPheLeuSerAs 640
 DB 1833 TTGGAGACTTCTTGAGCATACGATACAGACGCGACGTGCATACGCTGTACTCTCC 1892
 QY 640 n-----AsnAspLleThrGly-----GlyAsnSerGlySerProValPheAs 654
 DB 1893 TTATTTTCAGACATATATTCGCTGCTCCCTCATAGGCAATGCAAGATCT----- 1941
 QY 654 pLysAsnGlyArgLeuLleGlyLeuAlaPheAspGlyAsnTrrpGluAla 670

```

Db 1942 -----:::|||||
RESULT 2
AAK91621
ID AAK91621 standard; DNA; 2859 BP.
AC AAK91621;
XX
XX
XX 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG67 ORF encoding DNA.
KM Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX vaccine; antigenic; ds.
XX
XX Porphyromonas gingivalis.
OS
PN W09929870-A1.
XX
XX 17-JUN-1999.
PD
XX
XX 10-DEC-1998; 98WO-AU01023.
PF
XX
XX 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
XX
XX Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX WPI; 1999-385613/32.
DR P-PSDB; AAY34403.
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
XX Claim 12; Page 153-154; 588pp; English.
XX
XX AAK91536 to AAK91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAK91802 to AAK91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
XX Sequence 2859 BP; 738 A; 687 C; 692 G; 742 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.00725 Length: 2859
Score: 134.50 Matches: 105
Percent Similarity: 34.418 Conservative: 66
Best Local Similarity: 21.138 Mismatches: 202
Query Match: 3.628 Indels: 124
DB: 20 Gaps: 23
US-10-008-355-2 (1-712) x AAK91621 (1-2859)
QY 221 ArgHisThGlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArg-ProAl 240
DB 723 CGTCACAGGTCCGGTCATCTGTGTATCGCCGATATGCGCGATGATAGTCTTCTT 782

```

```

QY 240 aGIuYrSerIySaSpAsnIySProIyRlySProValTyrPheAlaAlaValSerMetI 260
DB 783 CCGCTATTTT-----CCTTTCCACAAATACCTTCGCTATATTCATGCC 830
QY 260 nGIyTyRlySaIaSpAspTyRAlaMetIhRlIleGIyPheProGIySerThraSparTy 280
DB 831 CACGTACGGAGAGACAAAT-----CGCTATGGATT-----TA 863
QY 280 rIeuthrSerTTPGIy-----ValGIuSparGIleGIuAsnGIuAsnBr 296
DB 864 TTTGAGGAATGGTGATATTATTTTCCTTCAGCGCATATACAT----- 909
QY 296 oArGIleGIuValArgGIy-----IleYSnGIyIleTyPlySgluAlaMetSerAl 314
DB 910 ----TTGGCATTCGCGGAGATCTTTTCCAAAGGCTCATGGGCGCATTTCCGCCCAATC 965
QY 314 aAspGIaIaThrArgIleYSyTyRAlaSerIyTyRAlaGInSerAlaAsnTyRly 334
DB 966 GAAATATTAAGAAAGGTATATAGTACACGGCTCTGC-----GAAGCCATTTAT----- 1014
QY 334 sAsnSerIleGIyMetAsnArgIyLeuAlaArgLeuAspValIleGIyArgIySargAl 354
DB 1014 ----- 1014
QY 354 aGIuGIuATgAlaPheAlaAspTTPleArgIySAsnGIyLySserAlaValTyRGIyAs 374
DB 1015 -----CTGCTATCGAAGTCCGGCGACAAATACGCTCCGGAGA 1052
QY 374 pValIeuser-----SerLeuGIuSAlaTyRlySgluGIyAlaIySAlaAs 390
DB 1053 CTACAGCAAGACACACAGCTGATATTCGATGAGACACACAGTCAAGATCCGAAGCCAA 1112
QY 390 nArgGIuMetIhTyRleuSerGIuThrLeuPheGIyGIyThrGIuValValArgPheAl 410
DB 1113 TCCT-----TTGCAACGTTGTGCGCCATGCAATTTTGCACCGGAG 1157
QY 410 aGIuPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla-----GIyIleuLy 428
DB 1158 CTATTTCCAGAAATGCGCTGATATACCACTATGATGCAATGACCCGCTACTGACACAGC 1217
QY 428 sSerIeuaAspAspIyTyRlySAspTyRleuProSerIeuaSparGIyValIeuProAl 448
DB 1218 AAGTTCCGCCGTGAGCTATTCGCGCAAGTTTCGGGTACTCTTTTCATTTACGGGTAG 1277
QY 448 aMetIeuaSpIle-----ValArgArgArgIleProAlaAspLyLeuProAspI 465
DB 1278 CATG---GATATCAGCCAGAACATGCGCATACGACGCGTACTTACCTTCCGATCT 1334
QY 465 ePheIySaenValIleAspLyS---LySPhelYSglYAspThrIySlyTyRAlaAspH 484
DB 1335 TTCGATTATATATGTCACCGCTTATTCCTTTCAAGCGGAAGACCCGTGAGACCGGAGG 1394
QY 484 eValPheAspLySserValaIaProTySerAspLySPhenIAlaMetIeulYSerSe 504
DB 1395 ATGATACGGAAGTTGAGTGTGGCTATTCGCGTACGTCCTCCATATGATCTTGACAA 1454
QY 504 tAspLySgluLySPhenAlaLySAlaIleGIuLySAspProAlaValGIuLeuSerIySe 524
DB 1455 AGAGAAAGATTGTCTCCAGAGCAATCTCGTGCGCATTTGAGAAATGATGATCGCATATTC 1514
QY 524 rValIleAlaAlaAlaArgAlaIleGIuAlaAspAlaMetAlaAsnAlaTyRAlaIleI 544
DB 1515 CGTACCGATCACTTGCATGCTGCTCTTTGTGGATTATTAATCTGACATATGGGGGTAA 1574
QY 544 uTySglYySArgLeuPheAlaGIyLeuArgIuMetTyR-----ProGIyAr 561
DB 1575 CTACATATAGTGTGTATACGAAGGCTATACGAAGTGTGTGATGAGATTAAGAAAC 1634
QY 561 gAlaIeuProSerAspAlaAsnPheThrMetArgMetSerTyRGIySerIleLySglYTy 581
DB 1635 ATTCTGCTTGGACACAGCACTATTAATTCGCGAGACTGTAC-----GATTA 1682

```

QY 561 rclubProGlnaspGlyAlaTPTPTyrAsnTyrHisThrThreGlyLysGlyValLeuGlnIuLy 601
 Db 1683 CAGCTGTGCGCAGC-----TTATCTACACATGTGTACGGTATGTTCAGCC 1730
 QY 601 sGlnaspProLysSerAspGlnPheAlaValGlnIuLsnIleLeuAspLeu----- 618
 Db 1731 TTGAAACCT-----TTTTCCTTCGAGCGACATCTCATATGATCCGTATCG 1778
 QY 619 -----PheArgThrLysAsnTyrGlyAr 626
 Db 1779 CTTACGCCCACTGTCAGTTCTCTATATGCGGACCTTCACGAAACCCGATATGGCTT 1838
 QY 626 gTyrAla-----GlnAsnGlyGlnLeuHisIleAlaPheLeuSerAs 640
 Db 1839 TTGGAGCTTCTTTCGACATACGATACGACAGCAAGCTGCTACGCTCTACTCTCC 1898
 QY 640 n-----AsnAspIleThrGly-----GlyAsnSerGlySerProValPheAs 654
 Db 1899 TTATTTCGACGATATTTCGGTCTCCCTCCATGCGCATGACGATCT----- 1947
 QY 654 pLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAla 670
 Db 1948 -----GTCAATTCTCTTTTGACAAACAACTTAGAGGCC 1980

RESULT 3
 AAS55647
 ID AAS55647 standard: DNA; 2652 BP.

XX AAS55647;
 XX AAS55647;

DT 13-FEB-2002 (first entry)

DE Streptococcus pneumoniae DNA for cellular proliferation protein #218.

XX Antisense: ds: prokaryotic cellular proliferation gene;

KW antibiotic; antibacterial; drug design.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

PN MO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELITR) ELITRA PHARM INC.

PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

DR WPI: 2001-611495/70.

DR P-PSDB: AAU37788.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

XX Claim 27; Seq ID No 9284; 511pp; English.

CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 2652 BP; 736 A; 609 C; 618 G; 689 T; 0 other;

Alignment Scores:

Pred. No.: 0.00913 Length: 2652

Score: 133.00 Matches: 106

Percent Similarity: 33.20% Conservative: 64

Best Local Similarity: 20.70% Mismatches: 164

Query Match: 3.58% Indels: 178

DB: Gaps: 25

US-10-008-355-2 (1-712) x AAS55647 (1-2652)
 QY 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
 Db 76 GATGTTTCAAGCCTTCAGCGCATCAAAAGCTATGCTTATTCATGCTTATTCACCA 135
 QY 204 ProSerSerValGlyLysPheGlyLysPheAspAsnTrpMetTrpProArgHisThr 223
 Db 136 CCAACGCTTACAGGTAACTT-----CACCTT 162
 QY 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTrpSer 243
 Db 163 GGT-----CACCT 171
 QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyLys 263
 Db 172 TGGGATACA---ACCTTCAGATATTTATTCATCCGTCATAAAGCATGCAAGGTTTGAT 228
 QY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTrpLeu---Thr 282
 Db 229 -----ACCCTTGGCTCTCTGAGGAGCACCAGCAGGATTCACACT 270
 QY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
 Db 271 CAGGCTAAGGTGAGAGCGCTGCGTGGAGGCAATTCGCCGTATGACCTTGCTGT 330
 QY 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
 Db 331 GAGCTTCTTGACGAAAGTCTGGGAATGAAAGC-----GAATATGCCACT 378
 QY 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
 Db 379 ACCATCAAGGAACAATGGGCGAGATGGCGCTCTGTAGCATATTCGTGAGCGTTTC 438
 QY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
 Db 439 ACTCTTGACGAAGGTTGTCA-----AAAGCTTTCGTAG 474
 QY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
 Db 475 GTCTTTGTGACCTTACAGAAAGCGTGAATCTACCGTGGATTTATCATCAACGTG 534
 QY 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLys----- 384
 Db 535 GACCAGCAGCTGCGACACCCCTTCTGATATGTAGTGCATCAACAGATGGAAGCT 594
 QY 385 -----GluGlyAlaLysAlaAsnArgGluMetThr 394
 Db 595 GCCTTTACACATGATTTACATGCTGTGAAAGATGGTTCACGCGCCCTTGAAGTGTCTACA 654

```

QY 395 TyrLeuSerGluThrLeuPheGlyGlyThrGluValValAlaPheAlaGlnPheAlaAsn 414
DB 655 ACTGCTCTAGACATGTTGGGGAGCTT----- 664
QY 415 AlaLeuAlaThrAsnProAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434
DB 685 GCGATTCGGGTAAATCCAGAA-----GACCGCGGTAC 717
QY 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
DB 718 AAGCACTGTATGTGTAATAAACGTCATCCTTCCA---ATCGCTAATAAACATCCATCCCATC 774
QY 449 MetLeuAsp-----IleValArgArg 455
DB 775 GTTGGAGATAGCAGCAGATCTGAGTTGTGACTGCTGCTGTAATAATCACACCTGCC 834
QY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
DB 835 CACGATCCAAATGACTCTTGGTTGGCCACGTCATACCTGCGCACACAGTCACATCATG 894
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
DB 895 AAGCAGACGCGAAGCTATGAGAGCTGTGCTTGAATTTTCAGGCATG-----GAT 945
QY 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
DB 946 CGTTTGAAGCTCGT-----AAGCAGACTCCTTCTTAAGTTGGAAGAATC 990
QY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaAlaAlaAlaAlaAla 531
DB 991 GGTGCCCTTGTCAAAATGCAAAAACGTGCCACAGTGTGTCACACAGAACGTCACAGT 1050
QY 532 -----IleGlnAla 534
DB 1051 GTCTGACTGAGCCAGCCTTGTCTACTCAATGCTGTCAGATGAGCAATTGGCTAAG 1110
QY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeu 554
DB 1111 AAGCCATTCGCAACAGACAGACAGAGACAG----- 1143
QY 555 ArgGluMetLysProGlyAlaAlaLeuProSer----- 565
DB 1144 GTGGAATTCACCCACCTCTTCACAGATACCTTCCTTCATGATGAGAAATGTCAC 1203
QY 566 AspAlaAsnPheThrMetLysMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
DB 1204 GACTGGGTATCTCTGCTCAGCTCTGCTGGGGGTAC-----CAATC 1245
QY 586 GlyAlaTyrPyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
DB 1246 CCGCTTGGTGAAT-----GCTGATGTGGAATATGATTCGCGCAAGAACGTCACGAA 1299
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
DB 1300 GGTGACGATGAGACTCAGAGCAAGACGTCCTGGAT 1335
RESULT 4
AAK57027
ID AAK57027 standard; DNA; 2652 BP.
XX
AC AAK57027;
XX
DT 19-JUL-1999 (first entry)
XX
DE S. pneumoniae vals coding region DNA.
XX
KM vals; treatment; antagonist; ss.
XX
OS Streptococcus pneumoniae.
XX
PN JP11113577-A.
XX
PD 27-APR-1999.

```

```

XX
PF 17-OCT-1997; 97JP-0321886.
XX
PR 17-OCT-1997; 97JP-0321886.
XX
PA (SMK ) SMITHKLINE BEECHAM CORP.
XX
DR WPI, 1999-320830/27.
DR P-SDB; AAY08339.
XX
PT New vals polypeptide and polynucleotide - useful for treatment of
PS conditions associated with abnormal vals protein levels
XX
PS Claim 4; Page 24; 27pp; Japanese.
XX
CC This invention describes the isolation of a novel Streptococcus
CC pneumoniae vals protein. The new polypeptide and antagonist are
CC useful for the treatment of an individual with abnormal vals
CC protein levels, by administering vals to individuals requiring
CC vals polypeptide, and administering the antagonist to individuals
CC requiring inhibition of vals polypeptide.
XX
SQ Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;
Alignment Scores:
pred. No.: 0.0177 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.81% Conservative: 62
Best Local Similarity: 20.70% Mismatches: 166
Query Match: 3.50% Indels: 178
DB: Gaps: 25
US-10-008-355-2 (1-712) x AAK57027 (1-2652)
QY 192 AspValPhe-----LysAspValArgMetValPheAlaPro 203
DB 76 GATGTTTCAAGCCTTCAGCGATCAAAAGCTAAGCTTATTCATCAATGTTATCCACCA 135
QY 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProAlaGlyHisThr 223
DB 136 CCAAACGTTACAGGTAAACTT----- 162
QY 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
DB 163 GGT-----CACGT 171
QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
DB 172 TGGGATGCA---ACTTGCAGATATATATCATCCGTCAAAACGATCAAGGTTTGTAT 228
QY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
DB 229 -----ACCTTTGGCTTCCTGCGATGAGCACGACGAGGATTCACAT 270
QY 283 SerTrpLysValGluAspArgIleGluAsnGlnAsnAsnProArgIleGluValArgGly 302
DB 271 CAGGCTAAGGTAGAGAGCCCTTCCGCTGAGGCACTTCCCGCTATGACTTGGTGCCT 330
QY 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
DB 331 GAGTCTTTTGTGACGAAGCTCGGAATGGAAGAC-----GAATATGCCACT 378
QY 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 357
DB 379 ACTATCAAGGACAAAGGCGCAAGATGCGCTCTGTATACATATCTCTGAGCGCTTTC 428
QY 338 GlyMetAsnArgGlyLeuAlaAlaArgLeuAspValIleGlyLysArgLysArgAlaGluArg 357
DB 439 ACTCTTGACCAAGTTGTCA-----AAAGCTGTTCCGAAG 474
QY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
DB 475 GTCTTGTGACCTTACAGAAAGCGCTGATCTACCGTGTGATGTTATCATCAACTGG 534

```



```

Db 229 -----ACCTTTGGCTTCCTGGGATGGACGACGAGGAGTTGCCACT 270
Qy 283 SerTrpGlyValGluSparGllEGluAnGluAsnSProArgllEGluValArgGly 302
Db 271 CAGGCTAAGGTAGAGCCCTTCGCTGGTGAGGCAATTTCCCGCATAGACCTTGCTGT 330
Qy 303 -----IleYsgInGlyIleTrpYsgLualMetSerAlaAspGlnAlaThr 318
Db 331 GAGCTTTCTTGACGAAGAAGCTCGGAATGGAAAGAC-----GAAATTCGCCT 378
Qy 319 ArgIleYsTrpAla-----SerLysTrpAlaGlnSerAlaAsnTrpTrpLysAsnSerIle 337
Db 379 ACATATCAAGACAACATGGGCGCAAGATGGGCTCTCTGTACACATCTCTCTGAGCGCTTC 438
Qy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
Db 439 ACCTTGACGACGAAGTTGTCA-----AAAGCTGTCGTAAAG 474
Qy 358 AlaPheAlaSprIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
Db 475 GCTCTTGAGACCTTTACAAAGAAAGGCTGGATCTACCGTGAGATTATCATCACTACG 534
Qy 374 Asp-----ValLeuSerSerLeuGluLysAlaTrpLysGluGlyAlaLys 388
Db 535 GACCCAGACAGCTCGACAGCCCTTTCTGATATTGAGGTATTCACAAGATGTGGAAAGT 594
Qy 389 AlaAsnArgIleMetIleTrpLeu----- 396
Db 595 GCCTTACACCATGATTCATGCTGGAAAGATGTTACACGCTCTTGAAGTTGTCTACA 654
Qy 397 -----SerGluTrpLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsn 414
Db 655 ACTCGTCCTAGACCAATGTTGGGAGCT----- 684
Qy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
Db 685 GCGGTTCCGGTCAACCCAGAA-----GACCCGCGCTAC 717
Qy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 718 AAGCACTTGATGTGTAATAATGTCATCTTCCA---ATCGCTAATAAACATCCCATCC 774
Qy 449 MetLeuAsp----- 455
Db 775 GTTGGAGATAGCAGCAGATCCTGAGTTGGTACTGGTCTGTAAGAAATCACACCTGCC 834
Qy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
Db 835 CAGATCCCAATGACTTCTTGTTGGCCACAGTCATTAATCTGCCACAAGTCACAGTCATG 894
Qy 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
Db 895 AAGCAGCAGCAACCATGATGACTTGGCTTGAATTTTCAAGGCATG-----GAC 945
Qy 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
Db 946 CGTTTGAAGCTCGT-----AAGCAGCTCTTCTAAGTTGGAAGAAATC 990
Qy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaIleAlaArgAla----- 531
Db 991 GGGGCCCTGTCATAAATCGAANAACGTCACAGTCTTGCTACCTCAGTCAGACCGTACAGT 1050
Qy 532 -----IleGlnAla 534
Db 1051 GTTGTGGTGAACCTGCTGTACTCAATGTTGCTCAAGATGACCAATATGGCTAAG 1110
Qy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
Db 1111 AAGGCCATTCACCAACCAAGACACAGACGACAG----- 1143
Qy 555 ArgGluMetLysTrpGlyArgAlaLeuProSer----- 565

```

```

Db 1144 GTCGAATTCACACACCTGCTTCAACGATACCTCTTCATGATGAGAAATGTCAC 1203
Qy 566 AspAlaAsnPheThrMetAlaGMetSerTyrGlySerIleLysGlyTrpGluProGlnAsp 585
Db 1204 GACTGGGTATTCCTCTGTCAGCTGCTGTTGGGTCAC-----CAATC 1245
Qy 586 GlyAlaTrpTyrAsnTrpHisThrTrpGlyLysGlyValLeuGluLysGlnAspProLys 605
Db 1246 CTGCTCCGTGACAT-----GCTGATGTTGAATGATGATTCGCGCAAGAACTCCAGAA 1299
Qy 606 SerAspLeuPheAlaValGlnGluAsnIleLeuAsp 617
Db 1300 GGTGACGAGATGACCTCAGACGAAGACGCTCTTGAT 1335

RESULT 6
AAAL4369
ID AAAL4369 standard; DNA; 2652 BP.
XX
AC AAAL4369;
XX
DT 15-AUG-2000 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae valyl tRNA synthetase (vals).
XX
KW Valyl tRNA synthetase; vals; inhibitor; stringent response;
KW drug screening; antibacterial; antibiotic; genetic immunisation;
KW antibody; bacterial infection; meningitis; ds.
XX
OS Streptococcus pneumoniae strain 0100993.
XX
FH Key Location/Qualifiers
FT CDS 1..1652
FT /tag "a
FT /product "Streptococcus pneumoniae valyl tRNA synthetase
FT (vals)"
XX
PD US6051413-A.
XX
PR 18-APR-2000.
XX
PF 24-SEP-1998; 9805-0159539.
XX
PR 18-APR-1996; 96GB-0007791.
PR 17-OCT-1997; 97US-0953492.
PR 18-APR-1997; 97US-0844064.
XX
PA (SMIR ) SMITHKLINE BEECHAM CORP.
XX
PI Brown JR, Lawlor EJ, Wang M, Jaworski DD;
XX
DR WPI; 2000-338311/29.
XX
DR P-PSDB; AAY90514.
XX
PT Novel vals polypeptides of valyl tRNA synthetase family useful for
PT treating otitis media, conjunctivitis, pneumonia and bacteremia
PT comprises a specified amino acid sequence -
XX
XX Claim 13; Columns 7-10; 21pp; English.
XX
XX
XX This sequence represents DNA encoding Streptococcus pneumoniae valyl tRNA
XX synthetase (vals; NCIMB deposit No. 40794). Streptococci are the cause of
XX several types of human diseases, including otitis media, conjunctivitis,
XX pneumonia, bacteraemia, sinusitis, pleural empyema, endocarditis and
XX especially meningitis. The frequency of S. pneumoniae infections has
XX risen dramatically over the past 20 years, probably due to the emergence
XX of multiply antibiotic resistant strains and an increasing population of
XX immunocompromised people. Vals represents a target for new antibacterial
XX agents. Inhibition of tRNA synthetases such as vals leads to a reduction
XX in the levels of charged tRNA, which triggers a cascade of responses
XX (known as the stringent response) resulting in a state of dormancy in the
XX bacterium. Vals, its variants and fragments, anti-vals antibodies, vals
XX inhibitors and nucleotides encoding vals may be used in the diagnosis,
XX prevention and treatment of bacterial infections such as meningitis.

```

CC Vals can be used to screen compounds for inhibitory activity. Vals
CC may also be useful as an antigen for vaccination of a host to produce
CC specific antibodies which protect against bacterial invasion into
CC damaged tissues. Such antibodies could, for example, prevent the
CC adherence of bacteria to wounds. Nucleotides encoding vals may be used as
CC diagnostic reagents and therapeutic or prophylactic agents, particularly
CC for genetic immunisation.

XX
SQ Sequence 2652 BP; 733 A; 612 C; 621 G; 686 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
130.00	2652	106	130.00
Percent Similarity:	Conservative:	62	32.81%
Best Local Similarity:	Mismatches:	166	20.70%
Query Match:	Indels:	178	3.50%
DB:	Gaps:	25	

US-10-008-355-2 (1-712) x AAAA14369 (1-2652)

```

Oy 192 AspvAlphe-----LysAspValArgMetValPheAlaPro 203
    |||||
Db 76 GATGTTTCAAGCCTTCAGCGCATCAAAAGCCTTATTCATATTCACCA 135
Oy 204 ProSerSerValGlyLysPheGlyLysPheAspThrPheAspThrProArgHisThr 223
    |||||
Db 136 CCAACGTTACAGGTAACCTT-----CACCTT 162
Oy 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGlySer 243
    |||
Db 163 GGT-----CACGT 171
Oy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
    |||||
Db 172 TGGGATAA---ACTTGCAGATATTTATCATCCCTCAAAACGATGCAAGTTTGAAT 228
Oy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
    |||||
Db 229 -----ACCTTGGCTCTCGGATGACACACGACGAGGATGCCACT 270
Oy 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGly 302
    |||||
Db 271 CAGGCTAAGATGATGAGGAGCGCTGCGTGTGAGGCAATTTCCCGCTATGACCTGTGCT 330
Oy 303 -----IleLysGlnGlyLeuTyrPheGlyLysGluAlaMetSerAlaAspGlnAlaThr 318
    |||||
Db 331 GAGCTTCTTGACGAAGTCTGGGATGGAAGAAC-----GAATATGCCACT 378
Oy 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTriPheAsnSerIle 337
    |||||
Db 379 ACTATCAAGGAACAATGGGGCAAGTGGGCTCTGTAGACTATTTCTGAGCGTTTC 438
Oy 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
    |||||
Db 439 ACTCTTGACGAAGGTTTGTCA-----AAAGCTTGTGTAAG 474
Oy 358 AlaPheAlaAspTyrIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
    |||||
Db 475 GTCTTTGTGACCTTACAAAGAAAGCTGATCTACCGGAGGATTAATCATCAACG 534
Oy 374 Asp-----ValLeuSerSerLeuGlyLysAlaTyrLysGluGlyAlaLys 388
    |||||
Db 535 GACCCAGACGCTCGACACGCCCTTCTGATATGTAGGTGATTCACAGAGATGTGAAG 594
Oy 389 AlaAsnArgGluMetThrTyrLeu----- 396
    |||
Db 595 GCCTTCTACACATGATTTACATGCTGGAAGATGTTACACGCGCTTGAAGTCTCA 654
Oy 397 -----SerGluThrLeuPheGlyLysThrGluValAlaArgPheAlaGlnPheAlaAsn 414
    |||||
Db 655 ACTGCTCTGAGACCATGTTTGGGACGTT----- 684
Oy 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434

```

```

Db 685 GCGGTTGGGATCAACCCAGAA-----GACCCGGGCTAC 717
Oy 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
    |||||
Db 718 AAGGACTTGATGTGTAATAATGATCATCTTCA---ATCGTAATTAACATCATCCAAATC 774
Oy 449 MetLeuAsp-----IleValArgArg 455
    |||
Db 775 GTTGAGATGAGCACGACGATCTGATTTGGTACTGGTGTCTGGAATAATCACACTGCC 834
Oy 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLys 475
    |||||
Db 835 CACGATCCAAATGACTTCTTGTGGCCACGATCACTTGCACAAAGTCACAGTCATG 894
Oy 476 GlyAspThrLysLysTyrAlaAspPheAlaPheAspLysSerValValProTyrSerAsp 495
    |||||
Db 895 AACGACGACGAGAACCATGATGACTTGGCTTTGAATTTTCAGGCATG-----GAC 945
Oy 496 LysPheHisAlaMetLeuLysSerMetAspLysGlyLysPheAlaLysAlaIleGlyLys 515
    |||||
Db 946 CGTTTGAAGCTCGT-----AAGCAGTCGTTGCTTAAGTTGGAAGAAATC 990
Oy 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
    |||||
Db 991 GGTGCCCTGTCAAAATGCAAAAGTGTCACAGTGTGGTCACTACAGACGTACAGAGT 1050
Oy 532 -----IleGlnAla 534
    |||
Db 1051 GTTGTGTTGAACCTCGCTTCTACTCAATGTTCTGCAAGATGACCAATTTGGCTAAG 1110
Oy 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPheAlaGlyLeu 554
    |||||
Db 1111 AACGCAATGTCACCAACCAAGACAGACAGACAG----- 1143
Oy 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
    |||||
Db 1144 GTCCAAATTCACCCACCTCGTTCAACGATACCTCTCAATGATGAAATGTCAC 1203
Oy 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
    |||||
Db 1204 GACGCGTTATCTCTCGGACACTCTGTGGGTGTCAC-----CAATTC 1245
Oy 586 GlyAlaTyrTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
    |||||
Db 1246 CTGCGCTGTGTAAT-----GCTGATGTGAATAATGATGTGCGGAGAAAGCTCCAGAA 1299
Oy 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
    |||||
Db 1300 GGTGACGGATGACTCAGACGAAGACGCTTGAT 1335

```

RESULT 7
AAV52160/c
ID AAV52160 standard; DNA; 6004 BP.
AC AAV52160;
AC AAV52160;
DT 23-OCT-1998 (first entry)
XX Streptococcus pneumoniae genome fragment SPQ ID NO:27.
DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX Streptococcus pneumoniae.
XX OS
XX W09818931-A2.
XX 07-MAY-1998.
XX 30-OCT-1997; 97WO-US19588.
XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX
 PI WPI; 1998-272225/24.
 XX
 DR Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 PS
 XX Claim 1; Page 305-309; 1409pp; English.
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC primers whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.
 XX
 SO Sequence 6004 BP; 1733 A; 1296 C; 1115 G; 1860 T; 0 other;

Alignment Scores:
 Pred. No.: 0.0661 Length: 6004
 Score: 129.00 Matches: 116
 Percent Similarity: 33.62% Conservative: 77
 Best Local Similarity: 20.21% Mismatches: 194
 Query Match: 3.47% Indels: 187
 DB: Gaps: 26

US-10-008-355-2.(1-712) x AAV52160 (1-6004)

QY 131 LysIIEValIysValThrsPlyValGIuGIyGlnLeuLysGIyIleThrsPgluMet 150
 DB 4315 AAATATAAATATATGCAATTTGTCGAGAAAGCTGCTAAGCAATTTCAAAATAAATA 4256
 QY 151 GIUAArgIleu-ArgIysAlaGlnGIuValIysGlnGIuLeuAlaLysGIuSnaIlaas 170
 DB 4255 AACAAATATAGAGAAAGACACACATGCTTAAGAATCTTCCATTAATACATCCAGC 4196
 QY 170 pGIuaSngIuLeuCySIIeValGIuProPheTySerAsnAsnGIuTyPheLeuIleVa 190
 DB 4195 CGAGGTTGAGGCTGTGCTGTACCAAAATG-----CTTGA 4160
 QY 190 ILYrAspValPhe-----LysAspValArgMetValPheal 202
 DB 4159 TGGCGATGTTTCAACGCTTCAAGCGATCAAAAGGCTAAACCTTAATCAATCGGTATCC 4100
 QY 202 APCProSerSerValGIyLysPheGlyGIyAspThrAspAsnTrpMetTrpProArgH 222
 DB 4099 ACCACCAAGTAACTGGGAACT-----CA 4073
 QY 222 sThrGIyAspPheSerValPheArgValIyAlaGlyAlaAspAsnArgProAlaGluTy 242
 DB 4072 CTTGGT-----CA 4064

QY 242 rSerIysAspAsnIysProTrIysProValIyTrPheAlaAlaValSerMetGlnIyTy 262
 DB 4063 CGCTTGGATACG---ACITTTGACGATATCATCCGCCAAAGCATGCAAGCTT 4007
 QY 262 rIysAlaAspAspTyAlaMetThrIleGlyPheProGlySerThrAspArgTyLeu-- 281
 DB 4006 TGAT-----ACCTTTGGCTTCCAGGATGAGCACCGCTGTATTCG 3965
 QY 282 -ThrSerTrpGIyValGIuAspArgIleGIuSngIuAsnAsnProArgIleGIuValAr 301
 DB 3964 GACTCAGCGTAAAGTTAGAGCGCTTACGCTGTGAGGATTTACCGCTTATGACCTAGG 3905
 QY 301 gGly-----IleLysGlnGIyIleTrpLysGIuAlaMetSerAlaAspGlnAl 317
 DB 3904 TCGTGAAGTCTTTCTTTCACGAGAGCTGCGAATGAAAGC-----CAATATGC 3857
 QY 317 aThrArgIleLysTyAla---SerIysTyAlaGlnSerAlaAsnTyTrpLysAsnSe 336
 DB 3856 CACTACCATTAAGCAACAAATGGGCAAGATGGGCTTCTGTGACTATTCTCGTGAGCG 3797
 QY 336 rIleGIyMetAsnArgGIyLeuAlaArgIleuAspValIleGIyArgLysArgAlaGlu 356
 DB 3796 TTTCACTGTTGATGAAGGTTTGTCA-----AAAGCTGTTCG 3761
 QY 356 uArgAlaPheAlaAspTrpIleArgLys-----AsnGIyLysSerValAlaTy 372
 DB 3760 TAAAGCTTTGTGACCTTTTACAGAAAGGCTGGATCTGCTGTGAGATTATTCATCA 3701
 QY 372 rGIyAsp-----ValIleuSerSerLeuGIuValAlaTyTrLys----- 384
 DB 3700 CTGGAGCCACGACGACGTCGACACCCCTTCTGTATTTAGAGCTATTACAAAGATGAG 3641
 QY 385 -----GluGIyAlaLysAlaAsnArgIuMe 393
 DB 3640 AGGTGCTTCTTACACATGATTTACATGCTGGAAGATGTTACGCGCTTGAAGTTGC 3581
 QY 393 tThrTyIleuSerGIuThrIleuPheGIyGIyThrGIuValValArgPheAlaGlnHeal 413
 DB 3580 TACAACTGCTGTCGAGACTATGTTGGGACGTT----- 3547
 QY 413 aAsnAlaLeuAlaThrsAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAsp 433
 DB 3546 ---GGGTTGGCGGTTAATCCAGAA-----GACCCGCG 3518
 QY 433 sTyTrIysAsp-----TyTrIeuProSerLeuAspArgLysValLeuPr 447
 DB 3517 CTACAAGAGCTTGATTTGTTAAAAACGTCAATCTTCCA---ATCGCTAATTAACATCC 3461
 QY 447 oAlaMetLeuAsp-----IleValAr 454
 DB 3460 AATCGTTGGAGATGAGACGACGACGATCCTGAGCTTGCTGCTGTAATAATCCACCC 3401
 QY 454 gATGATgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysSPH 474
 DB 3400 TGGCCACGATCCAAATGACCTTGTGGTGGCCAAAGCTGATTAATTCGCAACAATGCAAG 3341
 QY 474 eLysGIyAspThrLysLysTyAlaAspPheValPheAspLysSerValValProTySe 494
 DB 3340 CATGAACGACGACGAGAACTATGAATGAGCTTGCTTGAATTTTCAGGCAATG----- 3289
 QY 494 rAspLysPheHisAlaMetLeuLysSerMetAspLysGIuLysPheAlaLysAlaIleGI 514
 DB 3288 -GATCGTTTGAAGCTGT-----AAGGCACTGCTGTGAAGTGAAGA 3245
 QY 514 uLysAspProAlaValGIuLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
 DB 3244 AATCGGTGCCCTTGTCAAAATGCAAAACGTGCAACAGTGTGGCAGCTCAACAAGCTAC 3185
 QY 532 -----IleGI 533
 DB 3184 AGGTGCTGTAGTTGAGCCACGCTTGTCTACTCAATGCTTGTCAAGATGACCAATTTGGC 3125


```

QY 415 AlaleuAlaThrAspProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr 434
    |||:||||| |||:|||||
Db 685 GCGGTGGCGTTATTCAGAA-----GACCCCGGCTAC 717
QY 435 LysAsp-----TyrLeuProSerLeuSpArgLysValLeuProAla 448
    |||:||||| |||:|||||
Db 718 AAGCACTTGATGTGAATAAACGTCATCTCTCCA---ATCGCTATTAACATCATCCCAATC 774
QY 449 MetLeuAsp-----IleValArgArg 455
    |||:||||| |||:|||||
Db 775 GTTGGAGATGACGACGAGATCTGAGCTTGAGTGTGCTGCGTAATAATCATACACCTGCC 834
QY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
    |||:||||| |||:|||||
Db 835 CACGATCCAAATGACTCTTGTTGGTGGCCACAGCATATTAATCTTGCCACAAGTCACAGCTATG 894
QY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
    |||:||||| |||:|||||
Db 895 AACGACGACGGAATGATGATGAGCTTGCTTGATTTTCAGGCAATG-----GAT 945
QY 496 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
    |||:||||| |||:|||||
Db 946 CGTTTGAAGCTCGT-----AAGCAGCTCGTTGCTAAGTTGGAAGAAATC 990
QY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaIleArgAla-----IleGlnAla 531
    |||:||||| |||:|||||
Db 991 GGTGCCCTTGTCAAATATGAAAAACGTCATTCAGTCGCTCAGTCAAGAACGACAGCT 1050
QY 532 -----IleGlnAla 534
    |||:||||| |||:|||||
Db 1051 GTCGTAAGTTGACGACGCTTGCTACTCAATAGTTGCTGCAAGATGAGCAATTTGGCTAAG 1110
QY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
    |||:||||| |||:|||||
Db 1111 AACGCCATTTGGCAACAGACACAGACAGAGACAAG-----1143
QY 555 ArgGluMetTyrProGlyArgAlaLeuProSer-----565
    |||:||||| |||:|||||
Db 1144 GTCGAATTTTCACCACTCGTTTCAACAGATACCTTCTCATATGATGGAATAATGTCAC 1203
QY 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
    |||:||||| |||:|||||
Db 1204 GACTGGGTTATCTCTCGTCAAGCTTTGGGGGCTCAC-----CAATATC 1245
QY 586 GlyAlaIlePheTyrAsnTyrHisThrThyGlyLysGlyValLeuGluLysGlnAspProLys 605
    |||:||||| |||:|||||
*Db 1246 CCTCCCTGGTACAAAT-----GCTGATGATGTAATGTATGCGCGAAGAGCTCCAGAA 1299
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp 617
    |||:||||| |||:|||||
Db 1300 GGTGACGATGACTCAAGACGAAGACGCTCTTGAT 1335
RESULT 9
AAZ31948
ID AAZ31948 standard; DNA; 2142 BP.
XX
AC AAZ31948;
XX
DT 26-JAN-2000 (first entry)
XX
DE M. catarrhalis strain LES1 tbpb gene.
XX
KW TbpB gene; Tbp2; transferrin binding protein; diagnosis; otitis media;
KW genetic immunisation; Moraxella infection; antigen; vaccine; detection;
KW antitumour antibody production; therapy; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO952947-A2.
XX
PD 21-OCT-1999.
XX
PF 12-APR-1999; 99WO-CA00307.

```

```

XX
PR 14-APR-1998; 98US-0059584.
XX
PA (CONN-) CONNADUHT LAB LTD.
XX
PI Myers LE, Schryvers AB, Harkness RE, Loosmore SM, Du R, Yang Y;
PI Klein MH;
XX
DR WPI, 1999-620376/53.
DR P-PSDB; AAY43380.
XX
PT Nucleic acid encoding transferrin binding protein 2 of Moraxella
PT catarrhalis, useful for diagnostics, immunization and recombinant
PT protein production
XX
PS Claim 2; Fig 6; 114pp; English.
XX
CC This sequence encodes the Moraxella catarrhalis strain LES1 transferrin
CC binding protein (Tbp2) of the invention. This sequence is also referred
CC to as the tbpb gene. The tbpb gene is used to produce recombinant Tbp2;
CC for identification or diagnosis of Moraxella, or for cloning related
CC species, using hybridisation assays; and for genetic immunisation against
CC Moraxella infections, e.g. otitis media. The Tbp2 proteins are useful as
CC antigens, either in vaccines (including components of conjugate vaccines
CC that contain antigens from other bacteria or from tumours, in which case
CC they elicit production of antitumour antibodies that may be coupled to
CC chemotherapeutic agents or biologically active agents) or to raise
CC antibodies (for use as diagnostic reagents and for treating Moraxella
CC infections), also for detecting Moraxella antibodies.
XX
SQ Sequence 2142 BP; 753 A; 458 C; 457 G; 474 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.0561 Length: 2142
Score: 123.50 Matches: 125
Percent Similarity: 32.17% Conservative: 77
Best Local Similarity: 19.90% Mismatches: 273
Query Match: 3.32% Indels: 153
DB: gaps: 25
US-10-008-355-2 (1-712) x AAZ31948 (1-2142)
QY 119 LeuProIleProGlyLeuSerValLysTyrLeuArgLysIleValLysValThrAspLys 138
    |||:||||| |||:|||||
Db 466 ATTCCTTTTGTATAAAATCTTATGATTAACCTTAATAAAATCATCCGAGTTGTAAGTAA 525
QY 139 ValGluGlyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGlu 158
    |||:||||| |||:|||||
Db 526 TTTCAGACGCAAAAAGCGGATTTGAAATATACACAAGACTGACACACAAAAGATTATCA 585
QY 159 ValCysGlnGluLeuAlaLysLysGlnAsnAlaAspGluAsnGlnLeuCysIleValGlu 178
    |||:||||| |||:|||||
Db 586 TCAGAGCAAAAAGCAAAAAGTCAAAAGACGCTTGACAAATGCTTA-----633
QY 179 ProPheTyrSerAsnAsnGluTyrPheLeuIleVal-----TyrAspValPheLys 195
    |||:||||| |||:|||||
Db 634 ACTCAATTTGCCCAAGAAAATATACAAGAGCTATTTGCAAGCCCATGATATAAAATCT 693
QY 196 AspValArgMetValPheAlaProProSerSerValGlyLysPheGlyGlyAspThrAsp 215
    |||:||||| |||:|||||
Db 694 GACGCACGC-----702
QY 216 AsnThrPheThrProAlaGHisThrGlyAspPheSerValPheArgValTyrAlaGlyAla 235
    |||:||||| |||:|||||
Db 702 -----702
QY 236 AspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAla 255
    |||:||||| |||:|||||
Db 703 ---AACCGTATCTAGAAATATGATCAAGTCTGGTTTAACTATCTTCTGATATACCGCC 759
QY 256 Ala-----ValSerMetGlnGly-----TyrLysAla 264
    |||:||||| |||:|||||
Db 760 ACCGACGACGACAAAATAAATTAATATGCGCTATTAATGATGCTGATATATAAGGC 819

```

```

QY 265 AspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTyr 284
    ::: ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 820 AGCGAACCAGCCCAAGAGCTA-----CCACAACAAAGGCCAAATATTAAGGTTATTGG 873
QY 285 Gly-----ValGluAspArgIleGluAsnGluAsnAspProArgIleGluValArgGly 302
    ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 874 GACTTATGACAACTGCGACCTGATTAACAAATACACG-----GATTTGGCCAGCT 924
QY 303 Ile---LysGlnGlyIleTrrPlySGluAlaMetSerAlaAspGln---AlaThrArgGly 320
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 925 ATCGCCAGACAAACCCAGTGGCTTACTCTTGTCTTACTGATGAGTATGCACGCTCTTG 984
QY 321 -----LysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
    ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 985 ACAGACAAATAATACAAAGCCCAAGCTGATTAACATGTCATATGTCATGACAGTGAATTT 1044
QY 333 TrpLysAsnSerIleGly-----MetAsnArgGlyLeuAlaArg 345
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1045 GATGTTAATTTTGGCTGATTAATAAAATTAAGGCCAACTATCATGATCATGATATCAGCG 1104
QY 346 LeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLys 365
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1105 ACAGCTGTACCCGCCCAAGAGCTTATTAATAAGAACTGATATCCACGCGCAACCGCTTC 1164
QY 366 AsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGlu 385
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1165 CGTGGCAGTGCC-----ACCGCAAGCATTAACGA-----GAA 1197
QY 386 GlyAlaLysAlaAsnArgLueThrTyr---LeuSerGluThrLeuPheGlyGlyThr 404
    ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1198 GACAGCAAAACCCACACCCCTTACCGACGATGCTACACACGCTGAGAGTGGCTTTT 1257
QY 405 GluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAla 424
    ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1258 TATGACCAAAAGCGAGAGAGCTGCGAGTAATTTTAAACGATGACACAAACTCTTT 1317
QY 425 GlyIleLeuLysSerLeuAspArgLysTyrLysAspTyrLeuProSerLeuAspArgLys 444
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1318 GGGGTCTTGGTCTTAACGAGATTAAGTAAGAAACCGAACCCCTTAGATGCTAT 1377
QY 445 ValLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAsp 464
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1378 GCACCT-----GGG 1386
QY 465 IlePheLysAsnValIleAsp---LysLysPheLysGlyAspThrLysLysTyrAlaAsp 483
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1387 ACATTTAATATATCAATTAAGACACACATTCACCCCTTTACCAAAACAACTGGAT 1446
QY 484 ---PheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1447 AACTTTGGCAATGCCAAAAGTTGGTCTTGGTCTACGTCATTAATTTGGTGTCTAC 1506
QY 503 SerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSer 522
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1507 GATGCCACCAAAATGAATTCACCAAAATTCACCAAAAGAC----- 1548
QY 523 LysSerValIleAlaIleAlaIleArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 542
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1549 AAGCCAACTTGTGCCCAAAAGCGGGCGAGACT---TTGATGGTGAATGATGAAGTT 1605
QY 543 IleGluLys-----GlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGly 560
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1606 ATCGTCMAAAACGATGCGCAAAACTTTGAATACCTAAATTTGGTACGCTTACGTGCGT 1665
QY 561 ArgAla-----LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGly 576
    ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1666 GATAGCATAGCGCTTTTTCACAGCGCAACCGCTACACACAGGCGAGAAAGCGCGTA 1725
QY 577 SerIleLysGlyTyrGluProGlnAspGlyAlaTrrPtyAsnTyrHisThr----- 593
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1726 CCAACCACAGGCAAAAGCAATATCTGGGAACTGGGTAGATACATCACAGAGCGGCG 1785

```

```

QY 594 ThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlu 613
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1786 ACAGAAAGAAAGCTTAAATGAGGCCCAAGAT-----ATTGCTGAT 1824
QY 614 AsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeu 633
    ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1825 TTTGACATTGACTTTGAGAGAAATACGTTAAAGCCAAATGACCAACCCAGCGCGACA 1884
QY 634 HisIleAlaPheLeuSerAsnAsnAspIleThrGly----- 645
    ||| ::| ::| ::| ::| ::| ::| ::| ::|
Db 1885 GATCCTGCTTTTAACTCAAAAGTGAAATTCAGACGCAATGCTGCGACGCAAGCCAC 1944
QY 646 -----GlyAsnSerGlySerProVal 652
    ::| ::| ::| ::| ::| ::| ::| ::|
Db 1945 ACCACCAAGCGAGCGACGAGGCTACAGATGATTTACAGAGTACAGGCAAAATCCATC 2004
QY 653 PheAspLysAsnGlyArgGluIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer 672
    ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 2005 GTTCATCGAAATATGCCGAATTTACTGGGCTTTTATGTCCAATTCAAACGAGATGGC 2064
QY 673 GlyAspIleGluPheGluProAsp 680
    ||| ::| ::| ::| ::| ::| ::| ::|
Db 2065 GGGTCATTTACACAGCATACCGAT 2088

RESULT 10
AA164839
ID AA164839 standard; cDNA; 3271 BP.
XX
AC AA164839;
XX
DT 04-DEC-2001 (first entry)
XX
DE XRN-100 coding sequence.
XX
KW XRN-100; homologous recombination exchange and exoribonuclease protein;
KW chromosomal disorder; chromosome breakage syndrome; cancer; inflammation;
KW developmental disorder; immunological disease; haemopathy; gene therapy;
KW ss.
XX
OS Unidentified.
XX
PN CN1303931-A.
XX
PD 18-JUL-2001.
XX
PF 22-OCT-1999; 99CN-0119819.
XX
PR 22-OCT-1999; 99CN-0119819.
XX
PS (SHAN-) SHANGHAI BORONG GENE DEV CO LTD.
PI Mao Y, Xie Y.
XX
DR WPI: 2001-566159/64.
XX
DR P-PSDB: AAG78650.
XX
PT New polypeptide-homologous recombination medium-chain exchange protein
PT and exoribonuclease protein-XRN-100 for treating tumors, inflammation,
PT immunological disease and blood disease -
XX
PS Claim 6; Page 16-18(Disclosure); 27p; Chinese.
XX
CC The present invention provides the protein and coding sequences of
CC homologous recombination medium-chain exchange and exoribonuclease
CC protein XRN-100. The sequences can be used in the treatment of
CC chromosomal disorders, chromosome breakage syndrome, developmental
CC disorders, cancer, inflammation, immunological diseases and haemopathy.
CC The present sequence is the coding sequence of the invention.
SQ Sequence 3271 BP; 983 A; 648 C; 752 G; 888 T; 0 other:

Alignment Scores: 0.239 Length: 3271
Pred. No.:

```

Score:	119.50	Matches:	103
Percent Similarity:	34.76%	Conservative:	67
Best Local Similarity:	21.06%	Mismatches:	159
Query Match:	3.21%	Indels:	161
Gaps:	22	Gaps:	28

US-10-008-355-2 (1-712) x AAI64839 (1-3271)

QY	217	TriMetTrpProArgHisIstHngIysAspRheSerValPheArgValIyAlaGlyAlaAsp	236
Db	1614	TGGATTATTCATTCCATTCTATATGACCACTTTGGCT-----TCAGACTTTGAAAGCAATGGCA	166
QY	237	AsnArgProAlaGluIyTrSerLysAspAsnLysProTyrLysProValYTrPheAlaIa	256
Db	1668	GACATGCCATCTGATTTTGGAAAGGATGACAAACGGTTAAACACTA-----	171
QY	257	ValSerMetGlnGlyIyTrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySer	276
Db	1716	-----GAAACAACCTTATGGGGGTA-----TTTCCAGCTGGCA	174
QY	277	ThrAspArgTyrIleuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnAsnPro	296
Db	1746	AGTGCTAATTTTCTACTCCA-----	176
QY	297	ArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGln	316
Db	1767	-----TCATGGCGGAAGCTCATGAGTATCTCGAT	179
QY	317	AlaThrArgIleLysTyr-----AlaSerLysTyr	326
Db	1797	TCGTGTATTAATTCGCTCTATCTCGAAGATTTTGCATTTGATTTGAAATGGGAAGAAATAT	185
QY	327	AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyIleAlaArgLeu	346
Db	1857	GCA-----TGG-----CAAGTGTCCTCTCTTG	188
QY	347	AspValIleGlyArgLysArg-----AlaGluGluArgAlaPheAlaAspTrpIle	363
Db	1881	CCATTCGTGATGATGACGAAGGCTACGAGCTGCCCTAGAAAGATATACCCAGACCTGCAT	194
QY	364	---ArgLysAsnGlyLysSerAlaValTyrGlyAspValLeu-----	376
Db	1941	CCAGAAAGACACAGAAACACGCGTTGGAGTGATGTCCTTAATTTGTGGGAAACATCAC	200
QY	377	---SerSerLeuGluLysAlaTyrLysGlnGlyAlaLysAlaAsnArgLysMet	393
Db	2001	CCATCCCATGACTTCATTTAAGACCTTACCAGACAGCTTCC-----	204
QY	394	ThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAla	413
Db	2043	-----ACAGAGCCACTGGAG-----	205
QY	414	AsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLys	433
Db	2058	-----GTACCCCTCGAAGCTATGTGCATGGCAT-----CAAGGAAG	209
QY	434	TyrLysAspTyrLeuProSerLeuAspArgLys---ValLeuProAlaMetLeuAspIle	452
Db	2094	TTT-----TCTTGGATGAGAAGAGCCATTCTTCCAGATCAAAATAGATATG	213
QY	453	ValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsn---ValIleAsp	471
Db	2139	-----TCTCCGTGTTCCATGCTTAAAGGATCTGCACAGAACACTGTAGTCAGT	218
QY	472	LysLysPheLysGluArgPheThrLysLysTyrAlaAspPheValPheAspLysSerValAl	491
Db	2187	ATTTAATTTTAA---GACCCACACACTTGCAGAAATTAACATTTTAAAGCTGTAAATGCTT	224
QY	492	ProTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLys	511
Db	2244	CCAGAGGACGAACAAAGCCAGCAGCATGTACAAACCTGTACATGCGGAAAAATCCAGCAAT	230
QY	512	AlaIleGluLysAspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla	531

CC The present sequence represents the complete 1.66-megabase pair genome


```
OY 629 -----GLUasnGlyGlnIleuHisIleAlaPheLeuSerAsnAspIle 643
DB 5929936 ATTATTAGAGCTACAGACAGTGGTCTTACAGATATCAAGATTTAAAGATTAACATTAT 592995
OY 643 eThrGlyGlyAsnSerglySer-----ProValPheAspIlyAsnGlyArgLeuIleG1 661
DB 5929996 AGACGATGCTAAACAGTTCAGTACACACGCTTGAACAAAGATATATATG 593055
OY 661 ILeuAlaPheAspGlyAsnTrp 668
DB 593056 AGTTAAGTTCTGTGGAGACTGG 593077

RESULT 12
AAT42063/C
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KM Genome; bacterium; Haemophilus influenzae; computer readable medium;
KM expression modulating fragment; regulation; gene expression; vector;
KM organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYO) UNITV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H.influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
XX
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:
Pred. No.: 2e+03 Length: 1830121
Score: 117.00 Matches: 147
Percent Similarity: 30.62% Conservative: 98
Best Local Similarity: 18.38% Mismatches: 206
Query Match: 3.15% Indels: 349
DB: 17 Gaps: 42
```

```
us-10-008-355-2 (1-712) x AAT42063 (1-1830121)
OY 170 AspGluAsnGlnLeuCys----- 175
DB 605223 GATGAAATTCATATTACTTACTGGGGTTAATATTATCTCTCTAGTACAGAAATTATTA 605164
OY 176 -----IleValGluProPheTySerAsnAsnGlyArgPheLeuIle 189
DB 605163 TCTTAATTAATAAATAATTTTCTCTCTTTCTTTTAAGAATCCGCTATTTTCTT 605107
OY 190 ValTyArgPvalPheLysAspValArgPheValPheAlaProProSerSerValGlyLys 209
DB 605106 -----AGAAATAAGGAATATTACTACACCTTAATAAG----- 605074
OY 210 PheGlyGlyAspThrAspAsnTrpMetTrpProAlaGHisThrGlyAspPheSerValPhe 229
DB 605073 -----GAAACCTGTATGATGAACAGACGCTGATTTTATTCAAATTATGTCGACC 605026
OY 220 ArgValTyArgAlaGlyAlaAspAsnArgProAla----- 240
DB 605025 AGTATTCCTTGCATTATAGTCAAAACCGCTTTTGGCATGCAAAAGAAAGCATGTGAT 604966
OY 241 -----GluTySerLys 244
DB 604965 TTACGCAATTGTGCAACAACAGATGTCATAGTTCTTGACCGATTTGACATATATATAA 604906
OY 245 AspAsnLysProTyLysProValTyPheAlaAlaValSerMet----- 259
DB 604905 GATGCACCAACCGATTAATTCGTTTACGTCGCGGCAAGCCTTATTCGTCACACAGCT 604846
OY 259 ----- 259
DB 604845 GCTGAAGTAAAAAATAGCGTATTTGGTACATTAATGGGATTTATCAAGATCAATT 604786
OY 260 -----GlnGlyTyLys 263
DB 604785 GCAGACTATCAAGCCGCAACAGCGTATTAAGAGGGTAAATCAACCTGCATGATGT 604726
OY 264 -----AlaAspAspTyAlaMet---ThrIleGly----- 272
DB 604725 TTAAATGGGATGAATTAAGAACTGGTACATTATGATCAACACAGAAATTAACTATGCTGC 604666
OY 273 -----PhePro----- 274
DB 604665 AATTATCTTGCAATGCCATCAACAACAGCTAAATTTCCAAATTGTAATTAACAACGTACTA 604606
OY 275 -----GlySerThrAspArgTyLeuThrSerTrpGlyValGluAspArg-----Ile 290
DB 604605 AAGCGGGGACAGAAAGAACCTTATTTACACCTTATGTCATTCAGAAATAACACTGTTT 604546
OY 291 GluAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGly----- 306
DB 604545 GATTAATCAAGGT-----AAACACATAAATAAATAATGCTTACATCGTTT 604498
OY 307 -----IleTrpLysGluAla---MetSerAlaAspGlnAlaThrArg 319
DB 604497 GTGCCACCGCAATTAATGTTGGTATTAAGCAAACTTCAAGCAAAAGTGAACCCGT 604438
OY 320 -----IleLysTyAlaSerLysTyAlaGlnSerAlaAsnTyTrpLysAsnSerIle 337
DB 604437 GATATTGTAAAAACGACCAAAAATATGTACTGAA----- 604402
OY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
DB 604402 ----- 604402
OY 356 AlaPheAlaAspTrpIleArgTyAsnGlySerAlaValTyGlyAspValLeuSer 377
DB 604401 -----ATGAAGAAAAAAGGG-----GCTGATATTTGCTT 604372
OY 378 SerLeu-----GluLysAlaTyLysGluGlyAlaLysAlaAsn--- 390
DB: 17 Gaps: 42
```

```

Db 604371 GCATTAGCTCAGACAGGCCACTGTGATGACCTTATACAGAGGTGCTGAACCTCTGCA 604312
QY 391 -----ArgGluMetThrTyrLeuSerGluIuHrLeuPheGlyTyrGluValVal 407
Db 604311 TTCTATTATTAGCGGATGTCCTCATATTGATCGGTTATTTTGTCTACTACACCGCTTTA 604252
QY 408 -----ArgPheAlaGlnPheAlaAsnAlaLeuAlaIhrAsn----- 419
Db 604251 TTCCCAAAATTAAGAATTTGGCAAAATACCAAAATGCGGACATGTAATAGTACGGTAAAA 604192
QY 420 -----ProAspAlaIhAsnGly-----IleLeuLysSer 429
Db 604191 GGCATTACCAAGAAAGTATGGCTGGTACTGGCCAAATATATACAGCTGTGCAATTTAGC 604132
QY 430 LeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMet 449
Db 604131 TTAAACAGAACCAAGGTAATGATGCTGACTTCTGGCAAAAGCTGTACTTCTGTCCTCAATT 604072
QY 450 LeuAspIleValArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnVal 469
Db 604071 TATGATATAGAAACCAAAAAGCACTTGCAAAATAATGACCAAGAAATT-----ACGCGCA 604018
QY 470 IleAspLysLysPheLysGlyAspThrLysTyrAlaAspPheValPheAspLysSer 489
Db 604017 CTTTAAACCAAGTTCACGAAGCAGACAGCTAAATATGTTTCTCAACCTATGCGCAAAAGCC 603958
QY 490 ValValProTyrSerAspLysPheIleAlaMetLeuLysSerMetAspLysGluLysPhe 509
Db 603957 -----ACAGACAATATGATAGTACTT----- 603934
QY 510 AlaLysAlaIleGluLysAspProAlaValGluLeuSerLysSer----- 524
Db 603933 ---GCATTATTACAAAGATGATCCAAACATTCAAATTTGAACCAACAAAGAACCGTAT 603877
QY 525 -----ValIleAlaIleAla 529
Db 603876 GTAGAGAAAGTTGCACCAAGTATTGACAGCAATGGCTGGCTTACCTATTATTAAGTGCAGT 603817
QY 530 ArgAlaIleGlnAlaAspAlaMetAlaAsn-----AlaTyrAla---IleGluLys 545
Db 603816 GCACCATTTAAAGCAGCGGCTGTAAATAATGACCAACGGGTATTACGAGTGAATATAA 603757
QY 546 GlyLysArgLeuPhe---PheAlaGlyLeuArgLysMetTyrProGly----- 560
Db 603756 GGTAAATTAACCTTCCGTAAAGCTGCAGACTTATATCTCTCAAAATACCTAGTGGTT 603697
QY 561 ---ArgAlaLeuProSerAspAlaAsnPheIhrMetAlaMetSerTyrGlySerIleLys 579
Db 603696 GTTAAAGCAGACGCGCAACAATTAAGAGATGTTAGATGATGCTGTATGTTTAAA 603637
QY 580 GlyTyrGluProGln----- 584
Db 603636 CAATTTGACCTTACAAAGTAAACCAACATCAATTCATCGATGGAGGTTTCCGCACT 603577
QY 585 -----AspGlyAlaIhrPtyrAsnTyrIhrIsthrTh----- 594
Db 603576 TATAACTTTGATGTGATTTGATGTGTCATATTATGATATACGATCAACCAACAGCTCGT 603517
QY 594 ----- 594
Db 603516 TACGATGGCGAATGTAATTAATTAACCAACATCGCTGTAGTGAATCTCACTTAT 603457
QY 595 ---GlyLysGlyValIleGluLysGlnAspProLysSerAspGluPheAlaValGlnGlu 613
Db 603456 CAAGCGCAAAACAGTT-----GATCCAAAGCA---GAATTT----- 603424
QY 614 AsnIleLeuAspLeuPheArgThrLysAsnTyrGlyAlaGlyTyrAlaGlu----- 629
Db 603423 -----TTGATTGACAGCAATCAATCATGCTGTACGCGCAATTAATCCACAGT 603376
QY 630 AsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAsp----- 642
Db 603375 ACTGGTATTAACAAATATTGTTTATGCTGCTGCTGATGAAGCCGCTCAAAATTTGGCTGAT 603316

```

```

QY 643 -----IleThrGlyGlyAsnSerGlySer-----ProValPheAspLysAsnGly 657
Db 603315 TATATTAAAGCAACCACTGAGCAAAAGAAAGATCACTTATATCCAAATGGCGATAAAAACTGG 603256
QY 658 ArgLeuIleGlyLeuAlaPheAspGlyAsnTyrPgluAlaMetSerGlyAspIleGluPhe 677
Db 603255 CGTTTGTGCGCTATCACA-----GTTAACGATTAATTA-----GATGTCGTTT 603211
QY 678 Glu-----ProAspLeuGln-----ArgThrIleSerValAspIleArgTyrValLeu 693
Db 603210 GAAACATGCGCCAAAGCAAGCAAGCGCAAAATTTATCCGAGAAAAAGCAATATAC----- 603157
QY 694 PheMetIle-AspLysTyrGlyGlnCysProArgLeuIleGlnLysLeuLysLeu 712
Db 603156 ---GCATTGAACAAAGTGGGTACAGATGAATAGCTTTTGCAGTGTATCAATGATTT 603102

RESULT 13
AAC81763
ID AAC81763 standard; DNA: 15424 BP.
XX
AC AAC81763:
XX
DT 23-FEB-2001 (first entry)
XX
DE Porcine lelystad virus MSV JA-142 attenuated version.
XX
KW Porcine reproductive and respiratory syndrome virus; PRRSV; vaccine; pig;
KM attenuated virus; ds.
XX
OS Lelystad virus.
XX
PN W0200065032-A1.
XX
PD 02-NOV-2000.
XX
PE 21-APR-2000; 2000WO-US10852.
XX
PR 22-APR-1999; 99US-0298110.
XX
PR 15-DEC-1999; 99US-0461879.
XX
PA (USDA ) US SEC OF AGRIC.
XX
XX Mengeling WL, Vorwald A, Lager K, Roof M, Burkhardt K, Gorcyca DE;
PI WPI: 2000-687328/67.
XX
DR WPI: 2000-687328/67.
XX
PT Passaging viruses to attenuation comprises maintaining virus in
PT logarithmic phase of replication throughout multiple cell culture
PT passages, useful for protection against atypical porcine reproductive
PT and respiratory syndrome virus -
XX
XX
PS Example 7; Page 56-61; 70pp; English.
XX
CC The present invention provides a novel method of producing attenuated
CC versions of viruses, using multiple passage through cell cultures and
CC involving the removal of samples of the virus prior to the induction of
CC cytopathic effects. The sample is then inoculated into the next cell
CC culture. The sequence of the naturally-occurring lelystad virus (also
CC known as the porcine reproductive and respiratory syndrome virus or
CC PRRSV) atypical strain JA-142 is provided, along with that of a modified
CC version of the virus. The modified version can be used in the vaccination
CC of pigs against the lelystad virus.
XX
SQ Sequence 15424 BP; 3380 A; 4083 C; 4008 G; 3953 T; 0 other;

Alignment Scores:
Pred. No.: 3.71 Length: 15424
Score: 116.50 Matches: 161
Percent Similarity: 33.12% Conservative: 100
Best local Similarity: 20.43% Mismatches: 311
Query Match: 3.13% Indels: 217
DB: 21 Gaps: 40

```

US-10-008-355-2 (1-712) x AAC81763 (1-15424)

QY 7 SerIleuLeuEncllyAlaAlaLeuLeuEncllyAlaSerGlyValAlaLysAlaAspLys 26
 Db 10035 TTTAAATGTGTGGGACAGCTTCATCATCGCCCGCTGGT-----GCTGGGAA 10085
 QY 27 GlyMetIrrPheLeuAsnGluLeu-----AsnGln 36
 Db 10086 ACATACTGGCTCTTCAACAGGCTCCAGATGGATGCCATTACACGCCAATCCACAG 10145
 QY 37 GluAsnLeuAspArgMetArgGluLeuGly-----PheThrLeuProLeuAspSer 53
 Db 10146 ACCATGCTCATATGATTTAGAGCTTTGGGACGCTGCCGTTCAACGTCACAGCAGATACG 10205
 QY 54 LeuTyrSerPheAspLysProSerIleAlaAsn-----AlaValValIlePheGly 71
 Db 10206 AGCGTCAATTCCTCCGCCCTCCGATACCGCCCTGGGTTGGCATCTTACGCCGGCGGT 10265
 QY 72 GlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis----- 88
 Db 10266 TGGTGTCTCCGCAAGATTCCTCTCGATGAAGCAGCCGATTTGTAATCATTGATGTC 10325
 QY 88 ----- 88
 Db 10326 TTGAGGCTTCTAGCAAACTACCCCTCACCTGTCTGGAGATTTCAAACAATCCACCCA 10385
 QY 89 -----HisCysGlyTyrGlyAlaIleGlnSerLys----- 99
 Db 10386 GTGGGTTTGTATTCATCTGCTATGTTTGTGATCATGCTCATGCTCACTGAGAAC 10445
 QY 100 -----ThrValAspHisAspTyrLeuArgAspGly 109
 Db 10446 ATCTGAGATTGTGGCAAGATATCTGTGAGGCCATTCACGCAATATAC-----AGGACAAA 10502
 QY 110 PheValSerArgThrMetGlyGlnGluLeuProIleProGlyLeuSerValLysTyrLeu 129
 Db 10503 CTGTATCCATG-----GTCAACAACAACCCGTGTAACCTACGTG 10541
 QY 130 ArgLysIleValLysVal-----ThrAspLysValGluGly 141
 Db 10542 GAAAAACCTGTCAAGATGGGCAAGCTTCACCCCTTACCACAGGAGCCACAGAGGAGCGC 10601
 QY 142 GluLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValCysGln 161
 Db 10602 -----GCCATCCAAATGACCTCCAGTCAAGGCCCACTTGAATGTGTACACTG 10652
 QY 162 GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTyr 181
 Db 10653 CATTTGCCACTTAAGATCTACTCAACAGCACAAGAGCCCTGTGTCT-----ATT 10703
 QY 182 SerAsnAsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPhe 201
 Db 10704 ACCAGGGCAGACATGCTGCTTGTGTATGACCACACAGCAGCACTGCAGACATGTTT 10763
 QY 202 AlaProProSer-----SerValGlyLysPheGlyGlyAspThr 214
 Db 10764 GATCTTCCTCGAAGGACACACCCGTCACCTGCTGTGACCGT-----GACGAG 10814
 QY 215 AspAsnIrrPheTyrProArgHisThrGlyAspPheSerValPheArgValTyrAlaGly 234
 Db 10815 CAGCTGATCTGCTAGATTAATAACAAAGATGACGCTTCTGCTAGGCTAGGCAAT 10874
 QY 235 AlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPhe 254
 Db 10875 GGGGATTAATTCAGAGGCC-----ACAGACAAGCGCTTGTAGATTCCTCCGGCC 10925
 QY 255 AlaAlaValSerMetGlnGlyTyrLysAla-----AspAspTyrAlaMetThrIleGly 272
 Db 10926 ATTTGTGCAATCTGGAAGGCTGAGCTCCCGCTCCCAAGTGCACACAACTTGGGA 10985
 QY 273 PheProGlySerThrAspArgTyrLeuThrSer----- 283

Db 10986 TTTTATTTCTCGCTGAT-----TTGACACAGTTTGTAACTCCCGGTAGAACTTGCA 11039
 QY 284 -----TrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArg 301
 Db 11040 CCCCACTGGCCCGT-----GTGACAAACCCCAACAAAGAAAG----- 11078
 QY 302 GlyIleLysGlnGlyIleTrrPlysGluAlaMetSerAlaAspGlnAlaThrArgIleLys 321
 Db 11079 -----TGGCCAGACCGGTTGGTGTGTAGCTTGGCCCGCCGTCATFAG 11120
 QY 321 sTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrrPlys----- 334
 Db 11121 TATAGCGCGCGTCATCGTGGCCGCTTACATGCTGAGCGCCCTCACTGTTTGGGCACC 11180
 QY 335 -----AsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGly 350
 Db 11181 CTTGGGGTTGTGTATCATCTATCTCACAAATTTTGTACAGGGGCGAGGCTCAATGCTTCGG 11240
 QY 350 YArgLysArg-----AlaGluGluArgAlaPheAlaAspTrrPleArgLysAsnGly 367
 Db 11241 GAGACAGTCTTCAGACACCGCGCAATTTAGAGTGAATGGCTGATCTCGATACCGG 11300
 QY 367 YLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluGlyAl 387
 Db 11301 GAGCGA-----GAAATGCTGAGTCCCTCCCAATGCTTTCATTTGCGAGCT 11347
 QY 387 aLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValVa 407
 Db 11348 CAAAGC-----ACTACCGTTGGAGAGTTCACCATGT 11380
 QY 407 1-----ArgPhe-----AlaGlnPheAlaAsnAlaLeuAlaThr 418
 Db 11381 CACCTCCAAATACCTTCGCGCTTCTTCCCAAGAAATCAGTCCGCTAGTCCGGGTTTC 11440
 QY 418 rAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspAspLysTyr----- 434
 Db 11441 AAGCCCCGGGAAAGCCGCAAAACAGTTCGACATATGATGTATCTCCAGATCT 11500
 QY 435 -LysAspTyrLeu-----ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleVa 453
 Db 11501 CGAAGCTTACCTCCACCCAGACAGCCAGTCCAGAGTCTGGAATGATGTGGACTTCAA 11560
 QY 453 LArgArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLys 473
 Db 11561 GGAAGTTCGACTATGCTCGGAAGCACAAGAGCCGCTATTTCACACTTGAAGCCGCCCA 11620
 QY 473 sPheLysGly-----AspThrLysLysTyrAlaAspPheValPheAspLysSerValValPr 492
 Db 11621 TTTCACTGTGTACAGCTTGCAGAGTATGCTGCTGATATC-----CGAGTTCG 11668
 QY 492 cTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAl 512
 Db 11669 TGTTAAC-----TCTAC 11680
 QY 512 aIleGluLysAspProAlaVal-----GluLeuSerLysSerValIleAlaAlaIleAla 530
 Db 11681 GGTGTATTGTGACCCCTGCATGGGCCCTTGTGCACAGAGAGTGTGCGGTCCAC 11740
 QY 530 galatIleGluAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuP 550
 Db 11741 TCATTTGGGAGCTGACCTGCATGCCCTTAT-----GATTACGGTCCCAAAATCAT 11794
 QY 550 ePheAlaGlyLeu-----ArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAs 568
 Db 11795 CTTGTCTAGTCATACCATGCTGAAATGCCCTGGGTCAAAATCTGGCGTGGCGGA 11854
 QY 568 nPheIrrMet-----ArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGl 586
 Db 11855 GTTCTGCTTGGAGATTCAGTGAAGTACAAACACACTGGGGGTTTGA---TCGGATAC 11911
 QY 586 yAlaTrrPyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSe 606
 Db 11912 AGCG---TATCTGTACGAGTTCACCGAAACGCT-----GAGGACTGGGAGGATTCACAA 11962


```
D 10986 TTTATATTCGCGCTGAT-----TTGACACAGCTTGTCTAACTCCCGGTAGAACCTTGCA 11039
QY 284 -----TTPGLYValGluAspArgIleGluAsnGluAsnProArgIleGluValArg 301
Db 11040 CCCCACTGGCCCGTGG-----GTGACAAACCCAGAAACATGATAAG----- 11078
QY 302 GLYIleGlnGlnIleTTPGLYSGLuAlaMetSerAlaSpGln-AlaThrArgIleY 321
Db 11079 -----TGGCCAGACCGGTGTGGTCTGACCTTGCCGCCCTGCATTAAG 11120
QY 321 STYrAlaSerLYSTYrAlaGlnSerAlaAsnTYrTPLYS----- 334
Db 11121 TATAGCCGCGCGCATCGGTGCGGTACATGTTGGGCCCTCAGNCTTTCGGGCACC 11180
QY 335 -----AsnSerIleGlyMetAsnAArgLYLeuAlaArgLeuAspValIleGI 350
Db 11181 CTTGGGGTTGTGTCACTATCTCACAATAATTTGTCAAGGGCAGGCTCAAAATGCTTCCG 11240
QY 350 YArgLYSArg-----AlaGluGluArgAlaPheAlaAspTYrPLeaArgLYSAsnGI 367
Db 11241 GAGACAGCTTTCAGACACCGGCGCAATTGAGTGAATTGCGGTGATGATGACCGG 11300
QY 367 YLYSserAlaValTYrGLYAspValLeuSerSerLeuGluLYSAlaTYrLYSGLuGlyAl 387
Db 11301 GAGCGA-----GAAATGTGCTGAGTCCCTCCCATGCTTTCATTGGCGACGT 11347
QY 387 ALYSAlaAsnAArgGluMetThrTYrLeuSerGIuThrLeuPheGlyGLYThrGluValA 407
Db 11348 CAAAGGC-----ArgPhe-----AlaGlnPheAlaAsnAlaLeuAlaThr 418
QY 407 I-----ArgPhe-----AlaGlnPheAlaAsnAlaLeuAlaThr 418
Db 11381 CACCTCCAAATACCTCCCGGCTTCTCCCAAGAAATCAGTGGCGGTACTCGGGCTTTC 11440
QY 418 rAsnProAspAlaHisAlaGlyIleLeuLYSserLeuAspAspLYSTYr----- 434
Db 11441 AAGCCCGGGGAAAGCCCAAAACAGATGTCACATTAACAGATGTGTATCCACAGATCT 11500
QY 435 -LYSAspTYrLeu---ProSerLeuAspArgLYSValLeuProAlaMetLeuAspIleVa 453
Db 11501 CGAAGCTTACCTCCACCAAGACCCAGTCCAAAGTGTGTTGGACTTCA 11560
QY 453 LArgArgArgIleProAlaAspLYSLeuProAspIlePheLYSAsnValIleAspLYS 473
Db 11561 GGAAGTCCAGCTAGTGTCTCGAAGACAGACAGCGCTATTTCACCTTGAAGCGCCCA 11620
QY 473 sPheLYSGLY---AspThrLYSSTYrAlaAspPheValPheAspLYSserValAlaPr 492
Db 11621 TTTCACTGTGTACAGCTTCAAGCTATGCTGCTGATCATC-----CGAGTTCC 11668
QY 492 cTYrSerAspLYSPheHisAlaMetLeuLYSserMetAspLYSGluLYSPheAlaLYSAl 512
Db 11669 TGTTAAC-----TCTAC 11680
QY 512 aIleGluLYSAspProAlaVal-----GluLeuSerLYSserValIleAlaAlaAlaAr 530
Db 11681 GGTGTATTGTGGACCCCGCATGGGGCCCTTGCCATTTCAGAAAGAACTTGTCCGCTCAC 11740
QY 530 gAlaIleGluAlaAspAlaMetAlaAsnAlaTYrAlaIleGluLYSGLYsArgLeuPh 550
Db 11741 TCAATTGGGAGCTGACCTCCAGTCAACCCCTTAT-----GATTACGCTCCCAAAATCAT 11794
QY 550 ePheAlaGlyLeu-----ArgGluMetTYrProGlyArgAlaLeuProSerAspAlaAs 568
Db 11795 CCTGTCTAGTGCATACCATGGTGAATAGCCCGCTGGGTACAAATCTGGCGTGGCGGA 11854
QY 568 nPheThrMet-----ArgMetSerTYrGlySerIleLYSGLYTYrGluProGlnAspGI 586
Db 11855 GTTCTGCTTGGAGATGCATGAGTACAAACACACTGGGGGTTTGA---TGGATAC 11911
QY 586 yAlaTPYrAsnTYrHisThrThrGlyLYSGLYValLeuGluLYSGLnAspProLYSse 606
Db 11912 AGCG---TATCTGTACAGATTCAACCGGAACGGT-----GAGACACTGGAGAGATTCAAA 11962
```

```
QY 606 rAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLYSAsnTYrGLYAr 626
Db 11963 TGATGGCTTTCGCGCGCCAG-----AAAGGAAATTTATTAAGC 12004
QY 626 gTYrAlaGluAsnGLYIleuHisIle-----AlaPh 637
Db 12005 CACTGGCACCACGATGAGTTCATTTTCCCGGGCCCTGTCATTGAAACCACTTAGG 12064
QY 637 eLeuSerAsnAsnAspIleThrGlyLYSAsnSerGlySerProValPheAspLYSAsnGI 657
Db 12065 CTGAATTGAATGAATGATG-----GGGTCCATGCAAAAGCCTCTTGACAAAATTTGG 12115
QY 657 YArgLeuIleGlyLeuAlaPhe 664
Db 12116 CCAACTTTTGTGATGCTTTC 12137

RESULT 15
AAC77776
ID AAC77776 standard; cDNA; 3598 BP.
XX
AC AAC77776;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:170.
XX
KW Human: cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vunerary; immunomodulator;
KW antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; noctropic;
KW vasotropic; antiporiatic; antidiabetic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
PN WO20005350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000MO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
WP1: 2000-587533/55.
XX
P-PSDB: AAB43567.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 748-749; 2352pp; English.
XX
AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vunerary; immunomodulator;
CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC noctropic; vasotropic; antiporiatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditons.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
```

CC immune cells, to treat disorders of haematopoietic cells, autoimmune
 CC disorders, allergic reactions, graft versus host disease and organ
 CC rejection, modulate haemostatic or thrombolytic activity, modulate
 CC inflammation, cancers, cardiovascular disorders, neurological disease and
 CC bacterial or viral infections. The peptides, nucleotides, antibodies,
 CC agonists and antagonists may be also be used in drug screens. AAC78449 to
 CC AAC78457 and AAB44240 represent sequences used in the exemplification of
 CC the present invention.

XX Sequence 3598 BP; 1135 A; 636 C; 793 G; 1020 T; 14 other;

Alignment Scores:

Pred. No.:	0.59	Length:	3598
Score:	116.00	Matches:	110
Percent Similarity:	33.80%	Conservative:	85
Best Local Similarity:	19.06%	Mismatches:	198
Query Match:	3.12%	Indels:	185
DB:	21	Gaps:	30

US-10-008-355-2 (1-712) x AAC77776 (1-3598)

```

QY 90 CysGLYTYrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGly 109
   |||||:|||||
DB 39 TGTGCTTTCGGTGAG-----CTCTCAGACCGA 65

QY 110 PheValSerArgThrMetGlyGlnGluLeu-----Pro 120
   |||||:|||||
DB 66 TTTCTAGCTCCGTCGGGAGAGGTGCAGAGTCGRCTGTCGACATGCGCGCCT 125

QY 121 IleProGlyLeuSerValTyrLeuArgTyrIleValTyrValThrAspTyrValGlu 140
   |||||:|||||
DB 126 CCACGCGCG-GCCGGAGAGAGCGGATCCCAAGTCCCAAGTCCCAAGTCCCAAGTCCCGC 184

QY 141 GlyIleLeuLeuGlyIleThrAspGlnMetGluArgLeuArgTyrValGlnGluValCys 160
   |||||:|||||
DB 185 GCTGAGTACAG-ATACCTGCTGAA--CAACTCTTAAGAGAGGCTTAAGAAAGACAA 238

QY 161 GlnGluLeu-----AlaTyrSylGluAsnAlaAspGlnAsnGlnLeuCysIle 176
   |||||:|||||
DB 239 CTTCAGACTTCCTCCACCTCCACCTCAACAGAGATCACAGATGAGAGAAATTA----- 292

QY 177 ValGluProPheTyrSerAsnGlnTyrPheLeuIleValTyrAspValPheLeuAsp 196
   |||||:|||||
DB 293 -----AATGATTTTAACTAAGAGAAAGAGAAAGACTTTTGAAGAT 331

QY 197 ValArgMetValPheAlaProPheSerSerValGlyTyrPheGlyGlyAspThrAspAsn 216
   |||||:|||||
DB 332 -----AATATAAGAAAAAAGACGACTGTGATTAGTAAC 364

QY 217 TrpMet-----TrpProArgHisThrGlyAspPheSerValPheArgValTyr 232
   |||||:|||||
DB 365 TGGATTAATATACGACCAATG----- 385

QY 233 AlaGlyAlaAspAsnArgProAlaGluTyrSerTyrAspAsnLysProTyrIysProVal 252
   |||||:|||||
DB 386 -----GAGAAAGCCCTAAAGAGAGATTCAGAGCCCTGATCCACA 424

QY 253 TyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspTyrAlaMetThrIleGly 272
   |||||:|||||
DB 425 TACGAGCGGCTTTAGATGA-----GACTAC----- 451

QY 273 PheProGlySerThrAspArgTyrLeuThrSerTyrPheGlyValGluAspArgIleGluAsn 292
   |||||:|||||
DB 452 -----CGAATATATTACACTGCGCTGAATATACGCAAAATGGAATG 493

QY 293 GluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTyrLysGluAlaMet 312
   |||||:|||||
DB 494 AAGAAATCGCAAGTCAACCATGCTCGA-----AATATCTGGAGCGGCGCAAT 541

QY 313 SerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerLysAsnTyr 332
   |||||:|||||
DB 542 -----ACAAGCTGCTCGAGTTAAATCAAGTTCTGTGATACAAATGACGTAC 586

```

```

QY 333 TrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLys 352
   |||||:|||||
DB 587 ATGAGAGAAATGTGTGGA-----AACGTTGCCGTCGCCCG 622

QY 353 ArgAlaGlnGluArgAlaPheAlaAspTyrIleArgLysAsnGlySerLysValTyr 372
   |||||:|||||
DB 623 CAGGTGTTTGACCGC-----TGATGAGTGGCAGCCGAGAGACCAACCTGG 670

QY 373 GlyAspValLeuSerSerLeuGluLysAlaTyrLysGlnGlyAlaLysAlaAsnArgGlu 392
   |||||:|||||
DB 671 CACTCCTACATC--ACCTTGAGCTGAGATCAAAAG-----GTGATCGGGCC 718

QY 393 MetThrTyrLeuSerGlnThrLeuPheGlyGlyThrGluVal-----ValArgPhe 409
   |||||:|||||
DB 719 CCGCACTATTATAGACGAKTTCCTCGTGACCCCTGATGTTAAGACTGATCAAGTAT 778

QY 410 AlaGlnPheAlaAsnAlaLeuAla----- 417
   |||||:|||||
DB 779 GCCGCTTTGAAGAAAAACATGCTTATTTGCCCATGACGGAAGTGTAGACAGACT 838

QY 418 -----ThrAsnProAspAlaHisAlaGlyIle--LeuLysSerLeu 430
   |||||:|||||
DB 839 GTGCAATTCCTTGGAGATGAACATATGATGACACCTTTATGTCCTTGCCAAAGTT 898

QY 431 AspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeu 450
   |||||:|||||
DB 899 GAAGAAATATCAAGAAAGAGTTT-----GAAAGGTACAGAGTATTTACAGATATGCCCTG 952

QY 451 AspIleValArgArgIleProAlaAspLysLeuProAspIlePheLysAsn----- 468
   |||||:|||||
DB 953 GAC-----AGAAATTTCAAAACAGATGCCCAAGAACTCTTTAAATATATATACC 1000

QY 469 ValIleAspLysPheLysGlyAspThrLysTyrAlaAspPheValPheAspLys 488
   |||||:|||||
DB 1001 ATCTTGAGAAACATTT--GCTATAGCGCGGATATGAAGATATCATGTGTGACAAA 1057

QY 489 SerValValProTyrSerAspLysPheHisAlaMetLeuLysSerMetAspLys--Glu 507
   |||||:|||||
DB 1058 CGGAGATTCAGATGACGAGAAAGTGAAGCCAGCAACCAATATATGATCATGCTTT 1117

QY 508 LysPheAlaLysAlaIleGlnLysAspProAlaValGlnLeuSerLysSerValIleAla 527
   |||||:|||||
DB 1118 GATTACTTGGCGCTGTGAGAAAGTGAAGTGAAGCTGAA----- 1156

QY 528 AlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGlnLysGlyLys 547
   |||||:|||||
DB 1157 GCGGTGAGAGAGATGTATGAAGGSCATGTGCCATGTCCACCATTACAGAGAAAGG 1216

QY 548 ArgLeuPhePheAlaGlyLeuArgGlnMetTyrProGlyArgAlaLeuProSerAspAla 567
   |||||:|||||
DB 1217 CAC-----TGGAAAGCGCTACATTTAT----- 1237

QY 568 AsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAla 587
   |||||:|||||
DB 1238 -----CTT 1240

QY 588 TrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGlnLysGlnAspProLysSerAsp 607
   |||||:|||||
DB 1241 TCGATCACTATGCACTGTATGAAAGAA-----TTGAGAGCAAGAGATCTCGAGAGGACA 1294

QY 608 GluPheAlaValGlnGluAsnLleLeuAspLeuPheArgThrLysAsnTyr 624
   |||||:|||||
DB 1295 AGACAGGTGATCAAGCTCT--TTGGAACATAATTCCTCACAAAAGTTTC 1342

```

Search completed: October 17, 2002, 23:17:53
 Job time : 1848 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:44:46 ; Search time 54 Seconds
(without alignments)
3238.723 Million cell updates/sec

Title: US-10-008-355-2

Perfect score: 3719

Sequence: 1 NQMKAKSLIGALLIGASG.....LFMDKMGCCPRLIOELKLI 712

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_plus_model -DEV=xlh
-Q=/cg2_1/USPTO.spool/US10008355/runat.16102002.121154.27717/app_query.fasta.1.903
-DB=Issued_Patents_NA -QFMT=fastp -SUFFIX=tnl -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -CUFFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10008355.ecgn1.1.18.etrunc.16102002.121154.27717 -NCP=6 -ICPU=3
-NO_XLPHY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA:*
1: /cg2_6/ptodata/2/lna/5A.COMB.seq:*
2: /cg2_6/ptodata/2/lna/5B.COMB.seq:*
3: /cg2_6/ptodata/2/lna/6A.COMB.seq:*
4: /cg2_6/ptodata/2/lna/6B.COMB.seq:*
5: /cg2_6/ptodata/2/lna/PCTUS.COMB.seq:*
6: /cg2_6/ptodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	3.5	2652	2	US-08-953-492-1
2	115	3.1	4853	2	US-08-793-824-1
3	111	3.0	441529	4	US-09-103-840A-1
4	110	3.0	3279	4	US-08-446-137B-1
5	109	2.9	4403765	4	US-09-103-840A-2
6	106	2.9	2472	3	US-08-335-844A-7
7	106	2.9	3358	3	US-08-335-844A-20
8	106	2.9	3698	4	US-09-327-536-1
9	105.5	2.8	1994	5	US-08-600-982-22
10	105.5	2.8	1994	5	PCT-US94-10261A-22
11	104	2.8	3731	4	US-08-811-583-1
12	102.5	2.8	6142	4	US-09-514-302-1

13	102	2.7	2582	2	US-08-816-105A-2
14	102	2.7	4310	3	US-09-008-172-1
15	102	2.7	4310	4	US-09-210-361-5
16	101.5	2.7	2253	1	US-07-854-596B-39
17	101.5	2.7	3727	1	US-08-249-380-1
18	100.5	2.7	3278	1	US-08-484-105-13
19	100.5	2.7	3278	1	US-08-484-106-13
20	100.5	2.7	8501	4	US-09-298-367B-6
21	100	2.7	4330	3	US-09-310-293-1
22	100	2.7	4330	4	US-09-579-376-1
23	100	2.7	6744	1	US-08-119-125A-2
24	100	2.7	46899	1	US-08-471-119A-1
25	99.5	2.7	2167	3	US-08-861-083-83
26	99	2.7	3531	2	US-08-714-402-1
27	99	2.7	5433	3	US-09-157-021-35
28	99	2.7	5433	3	US-09-156-842-35
29	98.5	2.6	2012	4	US-08-026-143B-2
30	98.5	2.6	2012	4	PCT-US92-10621-2
31	98.5	2.6	2012	5	PCT-US94-02233-2
32	98.5	2.6	2217	1	US-08-712-241-4
33	98.5	2.6	5175	2	US-08-843-530B-3
34	98	2.6	2652	4	US-09-115-704-1
35	98	2.6	7253	4	US-09-268-347-35
36	98	2.6	7291	4	US-08-913-942-3
37	97.5	2.6	1404	3	US-08-961-083-9
38	96.5	2.6	1575	4	US-09-230-388-2
39	96.5	2.6	2136	2	US-08-949-941B-1
40	96.5	2.6	3728	1	US-08-111-939-1
41	96.5	2.6	4821	1	US-08-053-614-3
42	96.5	2.6	4821	1	US-08-316-397B-3
43	96.5	2.6	4821	2	US-09-034-306-3
44	96.5	2.6	4821	3	US-09-259-437-3
45	96.5	2.6	4821	5	PCT-US93-09782-3

ALIGNMENTS

RESULT 1
US-08-953-492-1
; Sequence 1, Application US/08953492
; Patent No. 5849555
; GENERAL INFORMATION:
; APPLICANT: Brown, James
; APPLICANT: Jaworski, Deborah
; APPLICANT: Lawlor, Elizabeth
; APPLICANT: Wang, Min
; TITLE OF INVENTION: NOVEL VALS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,492
; FILING DATE: 17-OCT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/844,064
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 9607991.8
; FILING DATE: 18-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gilmli, Edward R
; REGISTRATION NUMBER: 38,891

```

: REFERENCE/DOCKET NUMBER: P31458-4/1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-4478
: TELEFAX: 610-270-5090
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2652 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: US-08-953-492-1

Alignment Scores:
Pred. No.: 0.00287 Length: 2652
Score: 130.00 Matches: 106
Percent Similarity: 32.81% Conservative: 62
Best Local Similarity: 20.70% Mismatches: 166
Query Match: 3.50% Indels: 178
DB: Gaps: 25

US-10-008-355-2 (1-712) x US-08-953-492-1 (1-2652)

OY 132 AspValPhe-----LysAspValArgMetValPheAlaPro 203
DB 76 GATGTTTCAAGCCTTACAGCGATCAAAAGCGTATTCATCGTTATTCACCA 135
OY 204 ProSerSerValGlyLysPheGlyLysAspThrAspThrMetTrpProArgHisThr 223
DB 136 CCAACGTTACAGGTTAACTT-----CACCTT 162
OY 224 GlyAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSer 243
DB 163 GGT-----CACCT 171
OY 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
DB 172 TGGGATPACA--ACTTTCAGATATATTCATCCGTCAAAACGATGCAAGCTTTGAT 228
OY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu---Thr 282
DB 229 -----ACCTTTGGCTTCTCGGGATGGACCGACGAGGATGCCACT 270
OY 283 SerTrpGlyValGluAspArgIleGluAsnGluAsnProArgIleGluValArgGly 302
DB 271 CAGGCTAAGATGAGGAGCGCTTGGGTGAGGCAATTCCTCCGTATGACCTTGTCGT 330
OY 303 -----IleLysGlnGlyIleTrpLysGluAlaMetSerAlaAspGlnAlaThr 318
DB 331 GAGCTTTCTTGACGAAAGTCTGGAAATGGAAAGAC-----GAATATGCCACT 378
OY 319 ArgIleLysTyrAla---SerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIle 337
DB 379 ACTATCAAGACATGAGGCGCAAGATGGGCGCTCTGTAGACTATCTTCGTAGCGCTTTC 438
OY 338 GlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArg 357
DB 439 ACTCTTGACGAGGTTTGCA-----AAAGCTGTGCTAG 474
OY 358 AlaPheAlaAspTrpIleArgLys-----AsnGlyLysSerAlaValTyrGly 373
DB 475 GTCTTTGGACCTTTACAGAAAGCTGGATCTACCGTGGGACTTTATATCTCAACTGG 534
OY 374 Asp-----ValLeuSerSerLeuGluLysAlaTyrLysGluGluLysAlaLys 388
DB 535 GACCACGACGCTGCACAGCCCTTTCTGTATATGAGTGATTCACAGAGATGTGGAAGT 594
OY 389 AlaAsnArgGluMetThrTyrLeu----- 396
DB 595 GCCTTCTACACATGATATACATGCTGGAAGATGTTACACGGCTCCTTGAAGTGTCTACA 654
OY 397 -----SerGluThrLeuPheGlyLysThrGluValValArgPheAlaGlnPheAlaAsn 414

```

```

DB 655 ACTGCTCCTGAGACCATGTTGGGACGTT----- 684
OY 415 AlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434
DB 685 GCGTGTGGCGTCAACCCAGAA-----GACCCGGCTTAC 717
OY 435 LysAsp-----TyrLeuProSerLeuAspArgLysValLeuProAla 448
DB 718 AAGCACTTGATGTGAATAATGTCATCTTCCA--ATCGTAATTAATCATCCCAATC 774
OY 449 MetLeuAsp-----IleValArgArg 455
DB 775 GTTGGAGATGAGCAGCAGATCCTGAGTTGGTACTGGTGTGCGTAATAATCACACCTGCC 834
OY 456 ArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPheLys 475
DB 835 CACGATCCCAATGACTTTTGGTGGCCACAGCTCATTACTTGCCCAAGTCAAGTCAATG 894
OY 476 GlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSerAsp 495
DB 895 AACGACGACGAGACCATGATGACTTGGCCTTGAATTTTCAGGCATG-----GAC 945
OY 486 LysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGluLys 515
DB 946 CGTTTGAAGCTCGT-----AAGCGAGTGTGTTGCTAACTTGGAAGAAATTC 990
OY 516 AspProAlaValGluLeuSerLysSerValIleAlaAlaAlaArgAla----- 531
DB 991 GGTGCCCTTCGTCMAAATGCAAAACGTGTCCACAGTGTGGTCTACTGACAGCTACAGCT 1050
OY 532 -----IleGlnAla 534
DB 1051 GTTGTGTTGAACCTCGCTTGTCTACTCATGATGTTGTCGACAGATGACCAATTTGGCTAAG 1110
OY 535 AspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeu 554
DB 1111 AACGCCATTTGCCAACCAACACACAGACAGAGACAAG----- 1143
OY 555 ArgGluMetTyrProGlyArgAlaLeuProSer----- 565
DB 1144 GTCGAATTCACCCACCTCGTTTCAACAGTACTCTTCCTCAATGATGAGAAATGTCAC 1203
OY 566 AspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp 585
DB 1204 GACTGCTGTATCTCGACAGCTCTGTGGGTAC-----CAAAATC 1245
OY 586 GlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLys 605
DB 1246 CCTGCTGTACAAAT-----GCTGATGTTGAATGTATGTGGCGCAAGAAAGTCCAGAA 1299
OY 606 SerAspGluPheAlaValAlaGlnGluAsnIleLeuAsp 617
DB 1300 GGTGACGATGAGTACACAGACGAAAGCTCTTGGAT 1335

RESULT 2
US-08-793-824-1
: Sequence 1, Application US/08793824
: Patent No. 5981838
: GENERAL INFORMATION:
: APPLICANT: Simpson, Christine Lynn
: APPLICANT: Giffard, Phillip Morrison
: APPLICANT: Jacques, Nicholas Anthony
: TITLE OF INVENTION: Genetic Manipulation of Plants to
: TITLE OF INVENTION: Increase stored Carbohydrates
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESSES:
: ADDRESSER: Griffith Hack & Co
: STREET: Level 8, 168 Walker Street
: CITY: No. 5981838th Sydney
: STATE: New South Wales
: COUNTRY: Australia
: ZIP: 2060

```



```

1  Oy 436 spTyLeuProserLeuAspArgLyValLeuProAlaMetLeuAspLysPheValArgA 456
2      : : : : : ||| ||| ||| ||| : : : : :
3  Db 3335 CCTTGGAAATGACTTCCAAAGGCAAGTACGGGTGTCCTCTTATGAATGAAAGGCA 3394
4      : : : : : ||| ||| ||| ||| : : : : :
5  Oy 456 rgLleProAlaAspLysLeuProAspLlePheLysAsnValLleAspLysLysPheLysG 476
6      : : : : : ||| ||| ||| ||| : : : : :
7  Db 3395 AATACCCAGCA-----ATCTTGAGCGCGTCA-GAT----- 3425
8      : : : : : ||| ||| ||| ||| : : : : :
9  Oy 476 LyAspThrLysTyTAlaAspPheValPheAspLysSerValTyProTyrSerAspL 496
10      : : : : : ||| ||| ||| ||| : : : : :
11  Db 3426 -----TTCAAGGGCGGCTAAATTGACATCCAAATGAG 3456
12      : : : : : ||| ||| ||| ||| : : : : :
13  Oy 496 ySPheHISAlaMetLeuLys--serMetAspLysGluLysPheAlaLysAlaLleGlu 515
14      : : : : : ||| ||| ||| ||| : : : : :
15  Db 3457 AAATATCAGCAATGTCAGGCACAGTATTCATGAGACAAATATCCAAAGTACTGGAG-- 3514
16      : : : : : ||| ||| ||| ||| : : : : :
17  Oy 515 sAspProAlaValGluleuSerLysSerValLleAlaAlaAlaArgAlaLleGlnAlaAs 535
18      : : : : : ||| ||| ||| ||| : : : : :
19  Db 3515 -----CTCGCTAT-GTCTTACAGATTA 3536
20      : : : : : ||| ||| ||| ||| : : : : :
21  Oy 535 palMeAlaAsnAlaTyTAlaLleAlaLleGluLysGlyLysAlaGleuPhePheAlaGlyLeu 555
22      : : : : : ||| ||| ||| ||| : : : : :
23  Db 3537 CGCTACCAACCAATACTTCAGCGCTCAAGCAGGTCAAACTTCCTTCTTAAACAATATAC 3596
24      : : : : : ||| ||| ||| ||| : : : : :
25  Oy 555 gLlMetTyProGly-----ArgAlaLeuProserAspAlaAsnPheThrMetArgMe 573
26      : : : : : ||| ||| ||| ||| : : : : :
27  Db 3537 TGAATTACTACGGAAGTGGTTCCGTAGGGTGGATGAT-----GT 3638
28      : : : : : ||| ||| ||| ||| : : : : :
29  Oy 573 tSerTyGlySerLleLysGlyTyTyrGluPro-----GlnAspGlyAla-- 587
30      : : : : : ||| ||| ||| ||| : : : : :
31  Db 3639 CCAATACCTCTCAATAGTGGTCTACCTCTCTAAGATACCTTATTCAAAGTGGGCGCAA 3698
32      : : : : : ||| ||| ||| ||| : : : : :
33  Oy 588 -----TTPDyAsnThrLysThrLysGlyLysGlyValLeuGluLysGlnAspProLys 606
34      : : : : : ||| ||| ||| ||| : : : : :
35  Db 3699 CCAGTGGATATCTTGATATGAAGATGGCAACATGCTACAGGTGACGAGTTCATTGATG 3758
36      : : : : : ||| ||| ||| ||| : : : : :
37  Oy 606 rAspGluPheAlaValGlIngluLysLleuAspLysPheArgThrLysAsnTyGlyAr 626
38      : : : : : ||| ||| ||| ||| : : : : :
39  Db 3759 CAAGAAATCTTCTTCTTACAGCAATGAGTCTCCAGCTA-----CGTACATGCTCTTG 3809
40      : : : : : ||| ||| ||| ||| : : : : :
41  Oy 626 gTyTAlaGluAsnGlyLleuHisLle-----AlaPheLeuSe 639
42      : : : : : ||| ||| ||| ||| : : : : :
43  Db 3810 CCAAGTAGTAGTCATCTGATCTGATATTAACATCTTAAGGGGTTCAGGCCCTTAAACG 3869
44      : : : : : ||| ||| ||| ||| : : : : :
45  Oy 639 rAsnAsnAspLleThrGlyGlyLysAsnSerGlySerProValPheAspLysAsnGlyArgLe 659
46      : : : : : ||| ||| ||| ||| : : : : :
47  Db 3870 ATTTATGATATTTGGCGGCTCTCGCCACAGACGTTCCCTTATGATGGCAACGCTCAAT 3929
48      : : : : : ||| ||| ||| ||| : : : : :
49  Oy 659 u 659
50      : : : : : ||| ||| ||| ||| : : : : :
51  Db 3930 G 3930
52      : : : : : ||| ||| ||| ||| : : : : :
53  RESULT 3
54  DS-09-103-840A-1/c
55  : Sequence 1, Application US/09103840A
56  : Patent No. 6294328
57  : GENERAL INFORMATION:
58  : APPLICANT: FLEISCHMAN, Robert D.
59  : APPLICANT: WHITE, Owen R.
60  : APPLICANT: FRASER, Claire M.
61  : APPLICANT: VENTER, John C.
62  : TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
63  : TITLE OF INVENTION: TUBERCULOSIS
64  : FILE REFERENCE: 24366-20007.00
65  : CURRENT APPLICATION NUMBER: US/09/103,840A
66  : CURRENT FILING DATE: 1998-06-24
67  : NUMBER OF SEQ ID NOS: 2
68  : SOFTWARE: PatentIn ver. 2.1
69  : SEQ ID NO 1
70  : LENGTH: 4411529
71  : TYPE: DNA
72  : ORGANISM: Mycobacterium tuberculosis

```

OTHER INFORMATION: H37RV			
US-09-103-840A-1			
Alignment Scores:			
Pred. No.:	3.57e+03	Length:	4411529
Score:	111.00	Matches:	137
Percent Similarity:	29.45%	Conservative:	60
Best Local Similarity:	20.48%	Mismatches:	187
Query Match:	2.988	Indels:	286
Db:	4	Gaps:	35
US-10-008-355-2 (1-712) x US-09-103-840A-1 (1-4411529)			
QY	75	GIYLIETThrValSerAspBgInglYleuLlePheThrAsnHIScysGlyTYrGlyAla	94
Db	4023701	GGGCCACGGTG-----CAAGGCTTA-----CCGATATCGGGCAT-----	4023666
QY	95	ILeGInSerGlnSerThrValAspHisPtyrLeuArgAspGlyPheValSerArgThr	114
Db	4023665	-----GTCCGCACGGGGGTGCTCGCTTCGACATCTCGCGCGCATGGTTGCTCGGGCGT---	4023615
QY	115	MetGlyGluGluLeuProIleProGlyLeuSerValIstYrLeuArgGlyIleValIys	134
Db	4023614	-----GGCTACGACGCTCCGGCTTATCCCAACGTGACCGAC	4023579
QY	135	ValThrAspIlyValGluGlyGlnLeuIysGlyIleThrAspGluMetGluArgLeuArg	154
Db	4023578	ATCGAACAACAAGATC-----	4023564
QY	155	LysAlaGlnGluValCysGlnGlnLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeu	174
Db	4023563	-----CTGGCCAAAGGCCCGCGGGCGGC-----	4023540
QY	175	CysIleValGluProPheTYr-----SerAsnAsnGluTYrPheLeuIleVal	190
Db	4023539	-----CGGCCGTGTGGGAGTGGGGCGCATCCACGACGCTGCTTCACCGCGCC	4023489
QY	191	TyrAspValPheIlyAspValArgMetValPheAlaProProSerSerValGlyIysPhe	210
Db	4023488	TACAGACGCTG---GACGTC-----TTGCGCGCTGCCGGAG-----	4023453
QY	211	GlyIlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe---	229
Db	4023452	-----CCCGGCCGCCCGGGCATATACCGGACCAAGATGATC	4023420
QY	230	-----ArgValTYrAlaGlyAlaAspAsnArgPro	239
Db	4023419	GAGATGATCGAGCGCCGTGATCCAGCGCGCACCGCGATACCGGTGGCGGTAC-----	4023366
QY	240	AlaGluTYrSerIlyAspAsnIlyProTYrIlyProValTYrPheAlaAlaValSer---	258
Db	4023365	-----GTTTACTTCGACGTTCTGAGCTAC	4023342
QY	259	-----MetGlnGlyTYrIlySalaAspAspTYrAlaMetThrIleGlyPhe	273
Db	4023341	CCGGAGTACGGCCATTATCCGGTCCGACACAGATCATCATCATCAGCGCGAAGGTGTG	4023282
QY	274	---ProGlySerThrAsp---ArgTYrIleuThrSerTrpGlyValGlu--AspArgIleG	291
Db	4023281	GCTCCCGGCAAGCGGGACACAGCGGACTTCACTTTGGGAAGGGCGGAAGCCGGGTG--	4023224
QY	291	uAsnGluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpIlySGluAl	311
Db	4023223	-AACCGTCGTGGCGGACGCGGTGGGGCGGGGGGCTCGGGGCTGGCATTTGG-----	4023173
QY	311	ameTserAlaAspGlnAlaThrArgIleLysTYrAlaSerIlyTYrAlaGlnSerAlaAs	331
Db	4023172	-----AATGCTCG	4023165
QY	331	nTYrTrp-----LysAsnSerIleGlyMetAsnArgGlyLeuAlaAr	345
Db	4023164	GCAATGGCTCGCACATCTCTCGGGCGCGGAATTCGATTCGATTCGATTTGC--GTGGAAATGCATTT	4023106

US-08-446-137B-1

Alignment Scores:

Pred. No.:	0.0634	Length:	3279
Score:	110.00	Matches:	118
Percent Similarity:	31.59%	Conservative:	63
Best Local Similarity:	20.59%	Mismatches:	206
Query Match:	2.96%	Indels:	186
DB:	4	Gaps:	29

US-10-008-355-2 (1-712) x US-08-446-137B-1 (1-3279)

```

QY 29 TTPLeuLeuAsnGluLeuAsnGlnGluAsnLeuAspArgMetArgGluLeuGlyPheThr 48
    |||||
DB 1861 TGGTATTAAAGATCTTAAGACAAAGCAATTAAGAAATTAAGAACAGCAATCACT 1920
QY 49 LeuProLeuAspSerLeuThrLeuPheAspPylProSerLleAlaAsnAlaValIle 68
    |||||
DB 1921 -----TCTGATTTATAC--TTCAAGCTTAATCAATTAAGCAAAACAGTT----- 1962
QY 69 PheGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
    |||||
DB 1963 -----GAAGCGTAGAGCATTAAGACGAATTCCTTA-----AAAGCA 2001
QY 89 HisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeuArgAsp 108
    |||||
DB 2002 CACGCGTGA-----GAAGAAACACACAGAA-----TTAAAGAT 2034
QY 109 GlyPheValSerArgThrMetGlyGlnGluLeuProIleProGlyLeuSerValIlyStry 128
    |||||
DB 2035 GGTATATGCACATAGTAAAGACAGACAGCAAGCTTAAGAAAGCTTGAAAAATGATCAT 2094
QY 129 LeuArgLysIleValLysValThrAspLysValGluGlyGln-----LeuLys 144
    |||||
DB 2095 GTTAAACAGCATAGCAATAGTTCAAGTGCAGAGCAAGATACATGATATTAAAG 2154
QY 145 ----GlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValysGln----- 161
    |||||
DB 2155 ATTTGAAGTTGCACAGCAAGAAAGCAAGTGAAGACATCCAGCAAGTTCAAGAGGTAC 2214
QY 162 -----GluLeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCys 175
    |||||
DB 2215 GCACATTACAGAAAGACAGACAGCAAGCTTAAGAAAGCAATTAAGAAAGATTAAGTT--- 2271
QY 176 IleValGluProPheTyrSerAsnAsnGluTyrPheLeuIle----- 189
    |||||
DB 2272 -----AACATCGATACGAGAGTAGTACGATCAAGGTGACAGCGAAG 2310
QY 190 ValTyrAspValPheLys---AspValArgMetValPheAlaProProSerSerValGly 208
    |||||
DB 2311 TACTACTATATGATTAATAATCGAAGATTAAGAGATGAACACACAGTGAAGAACCGAGC 2370
QY 209 LysPheGlyGlyAspThr---AspAsnTrpMetTrpProArgHisThrGlyAspPheSer 227
    |||||
DB 2371 GAAACCCAGGAATCACAATTGATGATGTTTAAAGAAATGCTTAAGAAAGC----- 2424
QY 228 ValPheArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLys 247
    |||||
DB 2425 -----GCATTCAAAGAAATTAAGAAAGACGAGATCGTCTGAC----- 2463
QY 248 ProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLysAlaAspAspTyr 267
    |||||
DB 2464 -----ATATACCTTGTGCAATC----- 2481
QY 268 AlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSerTrpGlyValGlu 287
    |||||
DB 2482 -----AACAAAGCAAAACACAGTAGACGCGTAGAA 2511
QY 288 AspArgIleGluAsnGlu-----AsnAsnProArg 297
    |||||
DB 2512 ----GCGTTAAAGACGAATCTTAAAGACACAGCGTGAAGAAACCAAGCAAGCAAGCA 2568
QY 298 IleGluVal-----ArgGlyIleLysGlnGlyIleThrPylSgu----- 310

```

```

DB 2569 ATTCACATTTGATGAATGTTATTAAAGATGCTAAAGACAGTCGCAATCAAGCAATTAATA 2628
QY 311 -----AlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGln 328
    |||||
DB 2629 GAAGCAGGATACATCTGCTCAATTAATCTTAATCACTTAATCAACAAAGCAAGCAAGTAGAA 2688
QY 329 SerAlaAsnTyrTrpLysAsnSerIle-----GlyMetAsn 340
    |||||
DB 2689 GCGGTGATTCATTAAGAACGAATCTTAAGAACGACAGCTGAAACACAGCGCAAAAC 2748
QY 341 ArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGluGluArgAlaPheAla 360
    |||||
DB 2749 CCAGCAATC-----ACAAATTGAT 2766
QY 361 AspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGlu 380
    |||||
DB 2767 GAATGGTTATTAAAGAACCTTAA-----GAA 2793
QY 381 LysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400
    |||||
DB 2794 GATGCAATTAAGAA-----TTAAAGAACGAGGAATTAATCTTGACATATAC 2841
QY 401 PheGlyGlyThrGluValAlaArgPheAlaGlnPheAlaAsnAlaLeu----- 416
    |||||
DB 2842 TTTGATGCTATCAACAAAGCAAAACTATTGAAGCGGTGAAGCATTAAGCAATGAATAC 2901
QY 417 -----AlaThrAsnProAspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyr 434
    |||||
DB 2902 TTAAGGCTCATTAATAAGATGACAGACAGTAAATAAACCAAGTGAAGCAAAACCA 2961
QY 435 LysAspTyrLeuProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArg 454
    |||||
DB 2962 GAAGATTAATAAACACAGTGAAGTAAATA----- 2991
QY 455 ArgArgIleProAlaAspLysLeuProAspIlePheLysAsnValIleAspLysPhe 474
    |||||
DB 2992 -----CCAGAAAGACAAATAACCTGTTGAAGATTAATAAACCAAGCAAAACCA 3039
QY 475 LysGlyAspThrLysLysTyrAlaAspPheValPheAspLysSerValValProTyrSer 494
    |||||
DB 3040 CCAGTAAACAGATTA----- 3057
QY 495 AspLysPheHisAlaMetLeuLysSerMetAspLysGluLysPheAlaLysAlaIleGlu 514
    |||||
DB 3058 -----GATCTCCAAATTAAGAAAGCAAAAGCTTAATTA----- 3090
QY 515 LysAspProAlaValGluLeuSerLysSerValIleAlaAlaLargAlaIleGlnAla 534
    |||||
DB 3091 -----CCAAAGCTGTAGCGAAGCTGAATCTTAACATTAAGCAGACAGCAGCTTATCA 3144
QY 535 AspAlaMetAlaAsnAlaTyrAlaIleGlyLysGlyLys 547
    |||||
DB 3145 ACAGCAGCAGGCTTACGTTTCACCTTAATAAAACCTTAA 3183

```

RESULT 5

```

US-09-103-840A-2/C
: Sequence 2, Application US/09103840A
: Patent No. 6294328
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103.840A
: CURRENT FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 4403765

```

```

: TYPE: DNA
: ORGANISM: Mycobacterium tuberculosis
: FEATURE:
: OTHER INFORMATION: CDC 1551
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
: OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:      5 91e+03      Length:      4403765
Score:          109.00      Matches:      136
Percent Similarity: 29.34%      Conservative: 60
Best Local Similarity: 20.36%      Mismatches: 189
Query Match:      2.93%      Indels:      284
DB:               4          Gaps:      34

US-10-008-355-2 (1-712) x US-09-103-840A-2 (1-4403765)
QY 75 GlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHisScyGlyTyrGlyAla 94
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015782 GGTGCCACCGTC-----CAGGCGCTA-----CCGCATATCGCGCAT----- 4015747
QY 95 IleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGlyPheValSerArgThr 114
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4015746 -----GTCGCCAGCGGGGTCCTTCGACATCTGCCCGCATGGTGTCTCGCGGT---- 4015696
QY 115 MetGlyGlnGluLeuProIleProGlyLeuSerValLysTyrLeuArgGlyIleValLys 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4015695 -----GCGTACGACGTCGCGCTTATCCGCAACGTGACCGAC 4015660
QY 135 ValThrAspLysValGlnGlyGlnLeuLysGlyIleThrAspGlnMetGluArgLeuArg 154
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015659 ATCGAAGACAAATC----- 4015645
QY 155 LysAlaGlnGluValCysGlnGlnLeuAlaLysGlnAsnAlaAspGlnAsnGlnLeu 174
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015644 -----CTGGCCACAGCGCGCGCGCGGCGC----- 4015621
QY 175 CysIleValGluProPheTyr-----SerAsnAsnGlnLysTyrPheLeuIleVal 190
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015620 -----CGCGCGGGGGAGTGGCGGCTACCCAGACGTCGCTTACCCCGCGGCC 4015570
QY 191 TyrAspValPheLysAspValArgMetValPheAlaProProSerSerValGlyLysPhe 210
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015569 TACGACGCTCTG---GACGTC-----TTGCCCGCTCTCGCGAG----- 4015534
QY 211 GlyGlyAspThrAspAsnTrpMetTrpProArgHisThrGlyAspPheSerValPhe--- 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015533 -----CCGCGCGCCACCGCGCATATCCACCAGATGATC 4015501
QY 230 -----ArgValTyrAlaGlyAlaAspAsnArgPro 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4015500 GAGATGATCGAGCGCTGATCCAGCGCGGCCACGCGGTATACCGGTGGCGGTAC----- 4015447
QY 240 AlaGlnTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaAlaValSer--- 258
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015446 -----GTTTACTTGCAGCTTCTGAGCTAC 4015423
QY 259 -----MetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPhe 273
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015422 CCGGAGTACGCGCAGTTATCCGTCACAAGATCGATGATGCATCAGCGGCAAGGTGTG 4015363
QY 274 ---ProGlySerThrAsp---ArgTyrLeuThrSerThrGlyValGlu-AspArgIleG1 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4015362 GCTGCCGCGCAACGCGGACCGACGCTTCACTTTGTGAAAGGCGAAGAGCGGGGTG-- 4015305
QY 291 uAsnGlnAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrrPlysgluAl 311
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015304 -AACGCTGCTGGCGACGCGGTGGGGCGCGGCGCTCGGGGTGGATCTGG----- 4015254
QY 311 aMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAs 331
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015253 -----AATGCTCG 4015246

```

```

QY 331 nTyrTrp-----LysAsnSerIleGlyMetAsnArgGlyLeuAlaIar 345
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015245 GCATATGGCTCGCAGCTATCTCGCGCGGCAATTCGATTCATTCG- GGTGATGATTT 4015187
QY 345 gLeu-----AspValIleGlyArgLysArgAlaGlnGluArgAlaPheAl 360
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015186 ACTCTTCCCGCATCATGAGAAAGAAATCGCCAGAGCCGCGCGCGGTGATGGGTGCG 4015127
QY 360 a-----AspTrpIleArgLysAsnGlyLys-----SerAlaValTyr 372
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015126 CCGTACTGCTGCACAAATGCTGGGTGACGATGGCGGCGGAGAAAGATGCAATCGCT 4015067
QY 372 rGlyAspValLeuSer-----SerLeuGlnLysAlaTyr 383
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015066 GGGCAACGCTGTTCATCCAGCCGCGCATGTTGCACGCGGTGGCGCGCTCGACTCGTTA 4015007
QY 383 rLysGlnGlyAlaLysAlaAsnArgGlnMetThrTyrTyrLeuSerGlnThrLeuPheGlyG1 403
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4015006 TTAAGTGGGACGCGCGCATACCGGTGATGCTCGAGTTCTCGAGAC----- 4014958
QY 403 YThrGluValAlaArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaH1 423
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014958 ----- 4014958
QY 423 sAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAspAr 443
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014957 -----GCTATCAGAGATCGGTGAAGCCTTATGTGCGG----- 4014925
QY 443 gLysValLeuProAlaMetLeuAspIleValArgArgIle-----ProAl 459
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014924 -----CTAAGAGACTTCTCTCACCCTTGACACCGCGGTGGCGCGCTGCGCCCG 4014872
QY 459 aAspLysLeuProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLy 479
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014871 CGATCCAGCCCA----- 4014859
QY 479 sLysTyrAlaAspPheValPheAspLysSerValProTyrSerAspLysPheHisAl 499
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014859 ----- 4014859
QY 499 aMetLeuLysSerMetAspLysGlnLysPheAlaLysAlaIleGluLysAspProAlaYa 519
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014858 -----CGTTCGCCGAGCGCTCGACGACGCTGTGCT 4014824
QY 519 lGluLeuSerLysSerValIleAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAs 539
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014823 TCCGATCGCGCTCGCGGAGATTCACACACGTGGCGGCG----- 4014787
QY 539 nAlaTyrAlaIleGluLysGlyLysArgLeuPhePheAlaGlyLeuArgGlnMetTyrPr 559
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014786 -----GAAGGCAACCGGGCACTCGATGCGCGCGCACACGAC----- 4014751
QY 559 oGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLy 579
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014750 -----GGGCGCTTGCAGAGCGCTAGGCGCATGTGGGCGAGATG----- GGCATCTCT 4014704
QY 579 sGlyTyrGlnProGlnAspGlyAlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLe 599
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014703 GGGCTGTGACCCGCTCGACACGCTGTG----- 4014676
QY 599 uGluLysGlnAspProLysSerAspGlnPheAlaValGlnGlnAsnIleLeuAspLeu-- 618
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014675 -GAATCCGAGACGAACCTCGGACAGCGGTGGCGCGCGCTGATGTGCTGTCCAGGCTGA 4014617
QY 619 -PheArgThrLysAsnTyrLeuArgTyrAlaGlnAsnGlyGlnLeuHisIleAlaPheLe 638
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014616 ACTACAGAAATCGGAAGAGCCCGCGAGCAGCGCACTGTGGGGCTC----- 4014571
QY 638 uSerAsnAsnAspIleThrGly-----GlyAsnSerGlySerProValPheAspLysAs 656
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4014570 -----GCCAGCAGATTCGCGGTGCGCTGAAGAGACCGCATCGAGTCCACCGAC----- 4014520

```

OY 656 nglYargLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleG1 676
Db 4014519 -----ACCGCCGAGGGGCCACAGTGGCTGCTGGGTGTGACACCAA 4014476
OY 676 upheGluProAspLeuGlnArg 683
Db 4014475 GTAGATGCCGGAACCTCTCGG 4014454

RESULT 6
US-08-335-844A-7
; Sequence 7, Application US/08335844A
; Patent No. 6066503
; GENERAL INFORMATION:
; APPLICANT: GRAHAM, MARGARET
; APPLICANT: SMITH, TREVOR STANLEY
; APPLICANT: MUMN, EDWARD ALBERT
; APPLICANT: KNOX, DAVID PATRICK
; APPLICANT: OLIVER, JOANNA JANE
; APPLICANT: NEWTON, SUSAN ELIZABETH
; TITLE OF INVENTION: RECOMBINANT DNA MOLECULES ENCODING
; TITLE OF INVENTION: AMINOPEPTIDASE ENZYMES AND THEIR USE IN THE PREPARATION OF
; TITLE OF INVENTION: VACCINES AGAINST HELMINTH INFECTIONS
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz
; STREET: Suite 701-E, 555 Thirteenth St., N.W
; CITY: Washington
; STATE: D. C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/335,844A
; FILING DATE: 09-JAN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB93/00943
; FILING DATE: 06-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 920936
; FILING DATE: 08-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WALKER, Barbara W.
; REGISTRATION NUMBER: 35,400
; REFERENCE/DOCKET NUMBER: 1181-223A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)783-6040
; TELEFAX: (202)783-6031
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2472 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-335-844A-7

Alignment Scores:
Pred. No.: 0.113 Length: 2472
Score: 106.00 Matches: 111
Percent Similarity: 33.15% Conservative: 71
Best Local Similarity: 20.22% Mismatches: 230
Query Match: 2.85% Indels: 138
DB: 3 Gaps: 22
US-10-008-355-2 (1-712) x US-08-335-844A-7 (1-2472)

OY 89 HScYsGlyTYrGlyAlaIleGlnSerGlnSerThrValAspHisAspTYrLeuArgAsp 108
Db 283 CATCATGCTTCGCAATCTGTGTACACTGAAGTGGTGGATACCTGCTTGAACGAA 342
OY 109 GlnPheValSerArgThrmecGlyGluLeuProIleProGlyLeuSerValIysTYr 128
Db 343 GGTTTTGA-----ACATTGTTGATGATCTTGAAGAGCAAGAAATTAAGCCACAACAT 396
OY 129 LeuArgLysIleValLysValThrAspLysValGlnGlyLysLeuLysGlyIleThrAsp 148
Db 397 TTCACA-----ACCGAAGATTCTTCTTGGCTGCATGCAATGCAATCGCGGAATGAGACT 450
OY 149 GluMetGlu-----ArgLeuArgLysAlaGlnGluValCys 160
Db 451 GACGCGGAGCATCGAGCATCCGCTTGTGTTAGGATGACAAAGCCGAGAAAGTTGCC 510
OY 161 Gln-----Glu 162
Db 511 GAAGCCTTTGACGATATTTCATACGCCAAGGAGCGTCACTTCATATGCTACGGCT 570
OY 163 LeuAlaLysLysGluAsnAlaAspGluAsnGlnLeuCysIleValGluProPheTYrSer 182
Db 571 TTGATTGGAGAGACAAATTACAGAAATGCTGTTCGCAATACCTCAAGAAATTCCTAC 630
OY 183 AsnaGlnTYrPheLeuIleValTYrAspValPheLysAsp---ValArgMetValPhe 201
Db 631 AGCAATGCACAAGCAGCCGATCTGTGAACGCTTCAATGAATGAAGTGAAGTGAAG 690
OY 202 AlaProProSerSerValGlyLysPheGlyAspThrAspThrAspThrMetIleProArg 221
Db 691 GGTCTGACGGCAACGTCATGAAATTCAGCAATTTACCGACATGACGATAGATG 750
OY 222 HStrngLysPheSerValPheArgValTYrAlaGlyAlaAspAsnArgProAlaGlu 241
Db 751 GGTATCTCTGCTGCTTAAGTACAGAAATTTATATGACCCGCTTAAGCTTACGAGAGC 810
OY 242 TYrSerLysAspAsnLysProTYrLysProValTYrPheAlaValSerMetGlnGly 261
Db 811 CGGTACAGACCAATTAAGAC-----GCCTTGAAACCAAGAA 849
OY 262 TYrLysAlaAspAspTYrAlaMetThrIleGlyPhePro-----GlySer 276
Db 850 TATCGTAAATCCAAATACGGGTTCAAGTGGATGTTCCCTATGATCAGGAAGCAAT 909
OY 277 ThrAspArgTYrLeuThrSerTrpGlyValGluAspArg-----IleGluAsn 292
Db 910 ACCAAGAGCTAAACGCAACATGCTTAAGAAAGATGAACCCCTACTTGAACTCAAC 969
OY 293 GluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMet 312
Db 970 AATCGGATACATCCCTGTGTGTAACCTGATCGACATGATTTTATCGACA----- 1023
OY 313 SerAlaAspGlnAlaThrArgIleLysTYrAlaSerLysTYrAlaGlnSerAlaAsnTYr 332
Db 1024 -----AACTATGATCCCAACGCT 1041
OY 333 TrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLys 352
Db 1042 TCGAAAGA-----ATATCAAGCAGCTC 1065
OY 353 ArgAlaGlnGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTYr 372
Db 1066 AAGAAAGATCAACAGCTTCGTCACAGCAAG-----AACGCTATCATTA 1113
OY 373 GlyAspValLeuSerSer-----LeuGluLysAlaIleTYrLys--- 384
Db 1114 AGCGATGATTTGTCGACGCTACGATTGACCGAATGCACTATGAATCTATTCACACTA 1173
OY 385 ---GlnGlyAlaLysAlaAsnArgGluMetThrTYrLysSerGluThrLeuPheGlyGly 403
Db 1174 CTGGAATATGCCAAATAAGAGAGAAATTTCTTGGCTTGAAGAGAAAGCTGTGCCGCAATG 1233
OY 404 ThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHis 423


```

Db 1736 TATGTAATCAAAATAGGGTTCAAGTGGAGTTCCTCATGATGATCAGAGGCAAT 1795
QY 277 ThrAsparGlyLeuThrSerTrpGlyValGluAsparG-----IleGluAsn 292
Db 1796 AGCAAGAGGTGAAGCAATGCTGTAAGAGATGACCGCTGACTTGAACGTCAC 1855
QY 293 GluAsnAsnProArgIleGluValArgGlyIleLysGlnGlyIleTrpLysGluAlaMet 312
Db 1856 AATCGGATACATCCCTGTGTGTAACGCTGACATGATGATTTATGACAA----- 1909
QY 313 SerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
Db 1910 -----AACATGATGCCAACGGT 1927
QY 333 TrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLys 352
Db 1928 TGGAAAAAG-----ATAATCAAGCAAGCTC 1951
QY 353 ArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyr 372
Db 1952 AAGAAAGATCACAAGGCTTCGGTCCAGAGACAAG-----AACGCTATCATTA 1999
QY 373 GlyAspValLeuSerSer-----LeuGluLysAlaTyrLys--- 384
Db 2000 AGCGATGCAATTTCTGCTGCACTAGATGACGCAATGCACTATGAACTGTATTCGAACCTA 2059
QY 385 ---GluGlyValAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGly 403
Db 2060 CTGGAATATGCCAAAATGATGAGAGAAATTTCTGCTTGGAGAGAGAGCTGTCCGGCATG 2119
QY 404 ThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHis 423
Db 2120 TTCGAGTTTAAAGTTC-----TTCGTAATGAGCCGAGACAACAACA---GCTAGA 2170
QY 424 AlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAspArg 443
Db 2171 GCTTACATGATGACATATTATTAAGACCGAGTATTAATTAAG-----AGCAATGATGATTAC 2224
QY 444 LysValLeuProAlaMetLeuAspIleValArgArgGlyLeuProAlaAspLysLeuPro 463
Db 2225 ATGCTCAAGAAATTTTGGATGATACGTTTATCACAATAATTAATCTCAAAAG----- 2278
QY 464 AspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyrAlaAsp 483
Db 2279 GATATCATGATCATATTTGCTCCCTGATCAAGACAGTGTAAAGCAATATTAAGAT 2338
QY 484 PheValPheAspLysSerValAlaProTyrSerAspLysPheHisAlaMetLeuLysSer 503
Db 2339 ATCTTCTACGATAG--GTTATGCCAAGTGAAGCCGGGAGACGACCAACCAATGC 2395
QY 504 Met-----Asp 505
Db 2396 GTTAAGTTTCCGCTCTTCGAGCAATGTTTACTGTTATGTTGATACGAAAGGTGT 2455
QY 506 LysGluLysPheAlaLysAla-----IleGluLysAspProAlaValGluLeu 521
Db 2456 GAAGAAGCTTTGAAAAGGATGGGCTGTATCTAGCAGAAAGAT-----GTTCAACTG 2509
QY 522 SerLysSerValIleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAla 538
Db 2510 GAGAAAGGTATCTGTTCAAGCTTGGCATGCCACAAAGATGTATACAGCT 2560

```

```

; CURRENT FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: US 08/714,402
; PRIOR FILING DATE: 1996-09-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3698
; TYPE: DNA
; ORGANISM: SFBP gene
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (196)..(3681)
; US-09-327-536-1

Alignment Scores:
pred. No.: 0.211 Length: 3698
Score: 106.00 Matches: 90
Percent Similarity: 35.12% Conservative: 74
Best Local Similarity: 19.27% Mismatches: 140
Query Match: 2.85% Indels: 163
Gaps: 24

US-10-008-355-2 (1-712) x US-09-327-536-1 (1-3698)
QY 165 LysLysGluAsnAlaAspGluAsnGlnLeuCyIleValGluProPheTyrSerAsnAsn 184
Db 186 GAGAAATATTAATACACCAAAAATAGCTATAAGTTAAGCTTCTGTTATCCCTAACAG 245
QY 185 GluTyrPheLeuIleValIleTyrAspValPheLysAspValArgMetValPheAlaProPro 204
Db 246 ATTTATTTTAAAGTTTATTTATTTGCTTTTATAGCATTCGCG----- 287
QY 205 SerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProArgHisThr-Gl 224
Db 288 ---AGTATCATGAGCATGCCGA-----AACAGAAATGCGCAACAACAAG 335
QY 224 LysAspPheSerValPheArgValTyrAlaGlyAlaAspAsnArgPro----- 239
Db 336 ATCTTTTGAATACAGAAAGTGC-----GACCAAAACAAATTAAGCCTTACCGGAGCAAC 389
QY 240 ---AlaGluTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaValAla 258
Db 390 TTCTTACTGACATCAAGATGGCAAG-----GGAACATC 425
QY 258 TrpMetGlnGlyTyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAs 278
Db 426 TGTTCAAAGCTTACCTCAATGATTAAGTAT----- 459
QY 278 ParGlyTyrLeuThrSerTrpGlyValGluAspArgIleGluAsnGluAsnProArgTil 298
Db 460 -----GT 461
QY 298 eGluValArgGlyIleLysGlnGlyIleTrp-----LysGluAlaMetSerAlaAsp 316
Db 462 AGATGCTCAAAATCTCAACACGAGCTTATACCTTAAAGAAAGAAACAGACCAAT-- 519
QY 316 nAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsn 336
Db 520 -----GTTATGATAAACCACGCGGACTTGGACACTGC 554
QY 336 rIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArgLysArgAlaGlu 356
Db 555 TGTATTATGAGAAC--GCTATACCAAGTTG-----GTTCAAAA 590
QY 356 ValArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGlyAspVal 376
Db 591 TCCCTATTAATGGGAAATCATCATTAACAGAGGTCAAAA-----GATGTTAG 638
QY 376 uSerSerLeuGluLysAlaTyrLysGluGlyAlaLysAlaAsnArgGluMetThrTyr 396
Db 639 TACTTTTACAG-----TTGAAAAATCCCAAAATGTCACTGT 677
QY 396 uSerGluThrLeuPheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAla 416

```



```

Db 678 TTTAAA-----TATGGAAACAGAGGTT-----AGTATGGCGC 713
Qy 416 uAlaThrAsnProAspAlaHisAlaGlyIleLeuLys---SerLeuAspSplysTyRly 435
Db 714 AGCGATTTCACGCGAACACATGCGCCCTATTAAATGCTCTTTAGATTGAAACAAA 773
Qy 435 sAsp-----TyrLeuProSerLeuAspArgLysVa 445
Db 774 GGATTAATCTGAACAATCAACCCAGGTGATACCTTTGTGTACAGCTGATGATGACGCT 833
Qy 445 lLeuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuProAspII 465
Db 834 CAATCT----- 840
Qy 465 ePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyRAlaAspPheVa 485
Db 841 -----AAAGTATCATGCTCAAGATATCCCTTAATAATCAT 872
Qy 485 lPheAspLysSerValValProTyrSer---AspLysPheHisAla----- 499
Db 873 TTACGACAGTGCATAATAGTCCGCTTGGCATTTGGAATAATACCATCTGAGAACCATCACT 932
Qy 500 -----MetLeuLysSerMetAspLysGluLysPheAlaLysAlaII 513
Db 933 TATCTATACCTTTCACAGATTATATTGCGGGTTTATGATAAA----- 972
Qy 513 eGluLysAspProAlaValGluLeuSerLysSerVal-----lLeaAlaAlaIaAr 530
Db 973 -----GTCCAGTTGCTCGCAAGATTGAGCTTATCTCAAGATAAGAA 1016
Qy 530 gAlaIlleGlnAlaAspAlaMetAlaAsnAlaTyRAlaIlleGluLysGlyLysArgLeuPh 550
Db 1017 AGTGTGGAAATATCTGATCTCAATTTTAAGAGTACCATAGTGGCGGACGATCAC 1076
Qy 550 ePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPh-- 569
Db 1077 CTATTAAGGAACGGTATGATGTTCTTTATGGAATAGACAGCACTAAGAAAGCAATTATAT 1136
Qy 570 -ThMetArgMetSerTyr-----GlySerIleLysGlyTyRGlucProGlnAspGly-- 586
Db 1137 TACTAATGATGATGAGCATGTGGGTGGAGTATTAAGAAAGCTACAAACCGAAGCGGAGA 1196
Qy 587 -----AlaTrpTyrAsnTyr 591
Db 1197 ATTTGCTGTGATGTTAT 1215

RESULT 9
US-08-600-982-22
; Sequence 22, Application US/08600982
; Patent No. 6120991
; GENERAL INFORMATION:
; APPLICANT: Carter, William G.
; APPLICANT: G11, Susana A.
; APPLICANT: Ryan, Maureen C.
; TITLE OF INVENTION: Epiligrin, an Epithelial Ligand for
; TITLE OF INVENTION: Integrins
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson, and Kindness
; STREET: 1420 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-8100
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600, 982
; FILING DATE: 02-SEP-1994

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-8100
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: cDNA sequence corresponding to FIGURES
; DESCRIPTION: 11A-11C
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; US-08-600-982-22

Alignment Scores:
Pred. No.: 0.0915 Length: 1994
Score: 105.50 Matches: 112
Percent Similarity: 35.97% Conservative: 79
Best Local Similarity: 21.09% Mismatches: 201
Query Match: 2.84% Indels: 140
DB: 3 Gaps: 25

US-10-008-355-2 (1-712) x US-08-600-982-22 (1-1994)
Qy 69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
Db 514 TTCGAGAGTGTGCTCCACCATGCGATTGTAAACGCAATGCGACGCTGGCAGCTGCAT 573
Qy 89 HisCysGlyTyRgIyAlaIleGlnSerGlnSerThrValAspHisAsp----- 104
Db 574 CCCGTGACTGAGGCTGCATTA--AACCAAGAAACCAAGATTAACACCTGCGAAGAA 630
Qy 105 -----TyrLeuArgAspGlyPheValSerArgThr 114
Db 631 TGTGATGATTCGACACACCTGTGATGATGACCTCTCTGAAACAC-----CTGGCCACC 681
Qy 115 MetGlyGluGluLeu-----ProIleProGlyLeuSer----- 125
Db 682 ATGGCGGAGCAGCTCCGCTGTGCAAGTCTCAGCTGAGGCGCTGAGTGGCAGCGCAGGG 741
Qy 126 -----Val 126
Db 742 CTTCGTGAGCAGATGAGGACCATGAGACCCAGGCAAGGACCTGAGCATCACTGCTC 801
Qy 127 LysTyRLeuArgLysIleValLysValThrAspLysValGluGlyGlnLeuLysGlyIle 146
Db 802 AACCAACGTTCTCGCATTTAAATCATGATCAAAATAAGAGCCCGGAAGAGAACTG 861
Qy 147 ThrAsp-----GluMetGluArgLeuArgLysAlaGlnGluValLysGlnGluLeu 163
Db 862 ACTGATTGTAATCAAGAAATTGAGACTTTCAGAAAGAGCTCAAGTAAATTCACAGAAA 921
Qy 164 AlaLysLysGluAsnAlaAspGluAsnGlnLeuLysIleValGluProPheTyrSerAsn 183
Db 922 GCACAAACATTAAACACATGTTAATCGGCAACA-----CAAGCGCA 966
Qy 184 AsnGluTrpPheLeuIleValTyRAspValPheLysAspValArgMetValPheAlaPro 203
Db 967 AAAGAACTGATGAGAGATTAAATAAGTATCATCCGAATGCGACATTTCTTTAAACAG 1026
Qy 204 ProSerSerValGlyLysPheGlyGlyAspThrAspAsnTrpMetTrpProAlaGlnHisThr 223
Db 1027 ATCTCTGGACAGATGAGAGGAAACACAGCTGCT-----TCA 1065
Qy 224 GlyAspPheSerValPheArgValTyRAlaGlyAlaAspAsnArgProAlaGluTyrSer 243

```

```

Db 1066 GGTGACTTTTC-----AGAGAGTGGGCTGAAGCCAGCCGATGATGAGGAAGTCCGG 1119
QY 244 LysAspAsnLysProTyrLysProValTyrPheAlaIleValSerMetIleGlyTyrLys 263
Db 1120 AACAGCAAC-----TTTGGAAGCACCCTCAGAGAACGACGA 1155
QY 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSer 283
Db 1156 GGTGATAAAGGAGTCCGAGCTTGCTG-----AACGGATTAAGGACC 1200
QY 284 TrpGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIle 303
Db 1201 TGG-----CAGAAACCCACGAGGGGAGAACAAAT-----GGGCTT 1236
QY 304 LysGlnGlyIleTyrLys-----GluAlaMetSerAlaAspGlnAlaThr 318
Db 1237 GCTAACAGTATCCGGATTTCTTAATGATGATACGAAGCCAAACTCAGTACCTTCGTGCT 1296
QY 319 ArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTrpLysAsnSerIleGly 338
Db 1297 CGGCTGACGAGGACGCTGCCAAGCCAGACGCAAAAT-----GGC 1338
QY 339 MetAspArgGlyLeuAlaArg---LeuAspValIleGlyArgLysArgAlaGlu----- 355
Db 1339 TTGAAACCAAGAAACGAGAGAGCTTTGGAGCCATTCAAGACAACTGAAAGAAATTAAT 1398
QY 356 -----GluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSerAlaValTyrGly 373
Db 1399 TCCCTGCAGATGATTTCAACCAAGTATCTAACCCAGCTGACATCTTTGTTGCCAAAC 1458
QY 374 AspVal---LeuSerSerLeuGluLysAlaTyrLysGlu-----GlyAlaLys 388
Db 1459 AACATGCGCTGCAGCTGATGAGAAACCCAGAAAGATATGAAATAATTAAGCTGCCACT 1518
QY 389 AlaAsnArgGluMetThrTyrLeuSerGluThrLeuPheGlyGlyThrGluValValaG 408
Db 1519 TTAAATGAAGCAAGACACAGACTAGTGACAAAGTA-----AGAGAACTTTCCACA 1569
QY 409 PheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLys 428
Db 1570 TCTGCTGGC---AAACATCCCTTGTGGAGGAGCGCAAGAAACGACCG-----CGG 1617
QY 429 SerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 1618 TCCTTTCAAGAG-----CTGGCAAG 1638
QY 449 MetLeuAspIleValArgArgArgIleProAlaAspLysLeu----- 462
Db 1639 CAGCTGGAAGAGATCAAGAGAAAGCCAGCGGGGATGAGCTGTGCTGTGAT 1698
QY 463 ---ProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrLysLysTyr 481
Db 1699 GCCGCACCGCTCAGCAAGAAACATCCATGCCATCAAGCGCGCAGAGACGAGCAAC 1758
QY 482 AlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeu 501
Db 1759 AGGGCTCGCAGTGCATCTGATCTGCTGCT-CCAGACAGTGAATAAGGAAGATCTGCCAAG 1817
QY 502 LysSerMetAspLysGlu-----LysPheAlaLysAlaIleGlu 514
Db 1818 AAAAGCTAAACCTCGAGTTTCCACACAGTATAACTGTTTAAAGAACCAAGATGACACAA 1877
QY 515 LysAspProAlaValGluLeuSerLysSerVal 525
Db 1878 AAGAGCTAAAGCAAGAAAGTCAAGTCCAGCTCTC 1910

```

```

; TITLE OF INVENTION: Epiligrin, an Epithelial Ligand for
; TITLE OF INVENTION: Integrins
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: Christensen, O'Connor, Johnson, and Kindness
; STREET: 1420 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-8100
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10261A
; FILING DATE: 02-SEP-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-8100
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1994 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; DESCRIPTION: cDNA sequence corresponding to FIGURES 11A-11C
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; PCT-US94-10261A-22

Alignment Scores:
Pred. No.: 0.0915 Length: 1994
Score: 105.50 Matches: 112
Percent Similarity: 35.978 Conservative: 79
Best Local Similarity: 21.098 Mismatches: 201
Query Match: 2.848 Indels: 140
DB: 5 Gaps: 25

US-10-008-355-2 (1-712) x PCT-US94-10261A-22 (1-1994)
QY 69 PheGlyGlyGlyCysThrGlyIleThrValSerAspGlnGlyLeuIlePheThrAsnHis 88
Db 514 TTGGAGAGTACCTGCCAACCATGCTGTGTAAACAGCAATGGCCAGCTGGCAT 573
QY 89 HisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAspHisAsp----- 104
Db 574 CCCCTGACTGGAGCTGCATA--AACCAAGAAACCAAGATTAACAACTCCACAGAA 630
QY 105 -----TyrLeuArgAspGlyPheValSerArgThr 114
Db 631 TGTGATGATGGGACACCTGTGTGATGACCCCTCGAAGAC-----CTGGCACC 681
QY 115 MetGlyGluGluLeu-----ProIleProGlyLeuSer----- 125
Db 682 ATGGGCGAGCAAGCTCCGCTGCTCAAGTCTCAGCTGCAAGGCGCTGAGCCAGCGAGG 741
QY 126 -----Val 126
Db 742 CTTTGGACAGATGAGCAGCATGAGAGACCCAGGCCAAGGACCTGAGCATGTTGCTC 801
QY 127 LysTyrLeuArgLysIleValLysValThrAspLysValGlnGlyIleLeuLysGlyIle 146
Db 802 AACTACCGTTCGCAATTTCAATCATGATGATCAAAATAATGAAGGCGCTGGAAGAGAACTG 861
QY 147 ThrAsp-----GluMetGluArgLeuArgLysAlaGlnGluValCysGlnGluLeu 163

```

```

RESULT 10
PCT-US94-10261A-22
; Sequence 22, Application PC/TUS9410261A
; GENERAL INFORMATION:
; APPLICANT: Carter, William G.
; APPLICANT: Gil, Susanna A.
; APPLICANT: Ryan, Maureen C.

```

```

Db 862 ACTGATTGGAATCAAGATTGAGACTTTGCAAGAAAAGCCTCAAGTAAATTCAGAAAA 921
Qy 164 AlalysylsGluAsnAlaAspGluAsnGlnLeuGlyIleValGluProPheTyrSerAsn 183
Db 922 GCACAAACATTAAACAAATGTTAATCGGCAACA-----CAAAAGCGCA 966
Qy 184 AsnGluTyrPheLeuIleValTyrAspValPheLysAspValArgMetValPheAlaPro 203
Db 967 AAGAAGCTGGATGTGAAGATTAAAAATGTCAATCCGGAATGTGCACATCTTTTAAACAG 1026
Qy 204 ProSerSerValGlyLysPheGlyIleAspThrAspAsnTyrMetTyrProAlaHisThr 223
Db 1027 ATCTCTGGACAGATGAGAGAGAAACAGTGCCT-----TCA 1065
Qy 224 GlyAspPheSerValPheAlaGlyValTyrAlaGlyLysAspAsnArgProAlaGlyTyrSer 243
Db 1066 GGTGACTTTTCC-----AGAGACTGGGCTGAAACCCAGCGCATGATGAGGAACCTGCGG 1119
Qy 244 LysAspAsnLysProTyrLysProValTyrPheAlaAlaValSerMetGlnGlyTyrLys 263
Db 1120 AACAGGAAAC-----TTTGAAAGACCTCAGAGAAAGCA 1155
Qy 264 AlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeuThrSer 283
Db 1156 GCTGATTAAGAGGAGTGCAGCTCTTGCTG-----AACCGGATTAAGGACC 1200
Qy 284 TyrGlyValGluAspArgIleGluAsnGluAsnAsnProArgIleGluValArgGlyIle 303
Db 1201 TGG-----CAGAAACCCACACAGGGGAGAACAAAT-----GGGCTT 1236
Qy 304 LysGlnGlyIleTyrLys-----GluAlaMetSerAlaAspGlnAlaThr 318
Db 1237 GCATACAGTATCCGGGATTCTTTAATGAATACGAAAGCCAACTCACTGACCTTCGTGCT 1296
Qy 319 ArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyrTyrLysAsnSerIleGly 338
Db 1297 CGGCTGACAGGAGCGAGCTGCCCAAGCCAGCAGGCAAAAT-----GGC 1338
Qy 339 MetAsnArgGlyLeuAlaArg--LeuAspValIleGlyArgLysArgAlaGlu----- 355
Db 1339 TTGAACCAAGAAACGAGAGCTTTGGAGCCATTCAGACAAAGTGAAGAAATTAAT 1398
Qy 356 -----GluArgAlaPheAlaAspTyrPheArgLysAsnGlyLysSerAlaValTyrGly 373
Db 1399 TCCCTGCAGAGTATTCACCAAGTATCTAACCACTGCAGACTCATCTTTGTTCCAAACC 1458
Qy 374 AspVal--LeuSerSerLeuGluLysAlaTyrLysGlu-----GlyAlaLys 388
Db 1459 AACATTGCGCTGCAGCTGATGAGAAAGAAAGCAGAGAAATATGAATAATTAGCTGCCAGT 1518
Qy 389 AlaAsnArgGlyLeuMetThrTyrLeuSerGluThrLeuPheGlyIleThrGluValValArg 408
Db 1519 TTAATATAGACAGACAAAGACTTAAGTACAAAGTA-----AGAGAACTTTCAGA 1569
Qy 409 PheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAlaHisAlaGlyIleLeuLys 428
Db 1570 TCTGCTGGC---AAAACATCCCTTGTGAGAGAGCAGAAAGACGCGG-----CGG 1617
Qy 429 SerLeuAspAspLysTyrLysAspTyrLeuProSerLeuAspArgLysValLeuProAla 448
Db 1618 TCCTTACAGAG-----CTGCGCAAG 1638
Qy 449 MetLeuAspIleValAlaArgArgGlyIleProAlaAspLysLeu-----462
Db 1639 CAGCTGGAGATCAAGAGAAACCGCAGCGGATGAGCTGTGCGCTGTGTGAT 1698
Qy 463 ---ProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspThrIleLysTyr 481
Db 1699 GCCGCCACCGCCTACGAAACATCCTCAATGCCATCAAGAGCGCCGAGACCGGCAAC 1758
Qy 482 AlaAspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeu 501

```

```

Db 1759 AGGCGTCGAGTGCATCTGAATCTGCCCT--CCAGACAGTATTAAGAGATCTGCCAAG 1817
Qy 502 LysSerMetAspLysGlu-----LysPheAlaLysAlaIleGlu 514
Db 1818 AAAAGCTAAACCCCTGATTTCCAACTGATTAACCTTTAAAGAACCAAGATGACACAA 1877
Qy 515 LysAspProAlaValGluLeuSerLysSerVal 525
Db 1878 AAGAAGCTAAAGCAAGAAAGTACATCAGCTC 1910

RESULT 11
US-08-811-583-1
: Sequence 1, Application US/08811583
: Patent No. 6218142
: GENERAL INFORMATION:
: APPLICANT: Massenegger, Michael
: APPLICANT: Riedel, Leonhard
: APPLICANT: Schiebel, Winfried
: APPLICANT: Sanger, Heinz
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
: TITLE OF INVENTION: POLYPEPTIDES HAVING THE ENZYMATIC ACTIVITY OF AN
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: FISH & NEAVE
: STREET: 1251 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10020
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/811,583
: FILING DATE: 05-MAR-1997
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Haley, James F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: MPG-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-596-9000
: TELEFAX: 212-596-9090
: INFORMATION FOR SEO ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3731 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Tomato
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 194..3535
: US-08-811-583-1

Alignment Scores:
Pred. No.: 0.354 Length: 3731
Score: 104.00 Matches: 133
Percent Similarity: 32.06% Conservative: 77
Best Local Similarity: 20.31% Mismatches: 232
Query Match: 2.80% Indels: 213
DB: 4 Gaps: 31

US-10-008-355-2 (1-712) x US-08-811-583-1 (1-3731)
Qy 92 TyrGlyAlaIleGlnSerGlnSerThrValAspHisAspTyrLeuArgAspGlyPheVal 111
Db 1859 TATGCTGGAATATAAAGTGTGTCGGTGTGATCCGATTCATCAATGAAGATTGTCTTGG 1918

```

```

QY 112 SerArgThrMetCysGlu-----GluLeuProIleProGlyLeuSer 125
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1919 AGAAGAGCATGTCGAATAATGATACAGACAAACATAAATTTAGATCTCTTGATGACGAC 1978
QY 126 ValLysTyr-----LeuArgLysIleValLys 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1979 ---AAATATCAGCGCTGTATCTTAATCGTCAACGTATACGCTGTGCTACACTTGGGA 2035
QY 135 ValThrAspLysVal---GluGlyGlnLeuLysGlyIleThrAspGluMetGluValGlu 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2036 GTGAAGATGAGATGTTCTCGAAGCAGAGCAAAAGAAAGCTGTAGATCAGCTGTGCTATTC 2095
QY 154 -----ArgLysAlaGlnGluValCysGlnGluLeuAlaLysLysGlnAla 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2096 TTGCATGATCTTTTGAAGGCACAGAGGCTTTGGAAATGATGTCCTCGAGAGAACTACT 2155
QY 170 AspGluAsnGlnLeuCysIle-----ValGluProPheTyrSer 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2156 AATATTCTCAAGGCAATGCTAAACGTGGTTATAGCCGTAGCTGAGCCCTTTCTTTCA 2215
QY 183 AsnAsnGluTyrPheLeuIleValTyrAspValPheLys-----AspVal 197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2216 ---ATGATGTTGCAACACTTCGCGCATCCAAAGTTGCTCGATTGG 2257
QY 198 Arg-----MetValPheAlaProPheSerSerVal----- 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2258 CGGACTAGATCAAGAAATATTATTATTCCAAATGAGAAACAATGGATGGATTTGATGAA 2317
QY 208 -----GlyLys 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2318 TCCAGAACCTTGGAAATATGCTGACGTGTTGTTCACTTACTGCTGACATGAGAGAG 2377
QY 210 PheGlyGlyAspThrAspAsnIleMetCysProArgHisThrGlyAspPheSerValPhe 229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2378 TTTTCGAGAGATTACATCATTTATATACAGAGATCCACAGCATTAATTTCAATTCG 2437
QY 230 ArgValTyrAlaGlyAlaAspAsnArgProAlaGluTyrSerLysAspAsnLysProTyr 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2438 AAGGGAATATGCTGTTGTCAAAATAATCCATGCTGATCCTGATGATATGCTGTTTAA 2497
QY 250 LysProValTyrPheAlaIleValIleSerMetGlnGlyTyrLysAlaAspPyrAlaMet 269
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2498 AAGGCTGTAAATGTTGCGACGCTGCAC-----CACATGGTAGAT----- 2536
QY 270 ThrIleGlyPhePro-----GlySerThr 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2537 TGTGTTGATTTCCCTCAGAAAGAAAGAACCTCATCCAGATGAATGTTCTGGAGATGAT 2596
QY 278 -----AspArgTyrLeuThrSerTyrPglyValGluAspArgIleGluAsnGluAsn 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2597 TTGCATGGGATATCTACTTTGTTGCTCG-----GAT 2629
QY 295 AsnProArgIleGluValArgGlyIleLysGlnGlyIleTyrLysGluAlaMetSer--- 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2630 CAAGACATGATCCCGCCAGAGCAAGTCAGCCGATGAATATCTCCACACCCACACATTA 2689
QY 314 ---AlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyrAlaGlnSerAlaAsnTyr 332
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2690 CAGTTGGACCATGATGTCACAAATTTGAGAAAGTTGAAGAGATAC-----TTCACCACTAT 2743
QY 333 TyrLysAsn---SerIleGlyMetAsnArgGlyLeuAlaArgLeuAspValIleGlyArg 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2744 ATTGTGAATGACAGTTTGGGAATC-----ATAGCAAAAGCCCATGTGATTTTGA 2794
QY 352 LysArgAlaGlnGluValArgAlaPheAlaAspTyrPheArgLysAsnGly----- 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2795 GACAGAGAACTGATGTGCGACAGATGATCCATGCAAGAAAAAAGCTGCTAGACTTTTGA 2854
QY 368 -----LysSerAlaValTyrGlyAspValLeuSerSerLeuGlu--- 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2855 ATTGCAGTGACTTTCCAAAGAGCTGCTGTTCCCGCTGAATAATACATCATGATGGCCCT 2914

```

```

QY 381 LysAlaTyrLysGlnGluValAlaLysAlaAsnArgGluMetThrTyrLeuSerGluThrLeu 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2915 AAAAGATACCCAGACTTCATGATGAAGCGGAGACAGACAGCAGCTATATCTCAGAAAGACT 2974
QY 401 PheGlyGlyThrGluValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnPro 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2975 ATTTGA----- 2980
QY 421 AspAlaHisAlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSer 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2981 -----AAGCTTTTCAGGAAAGTGAAGGACAAACACTCTGAG 3016
QY 441 LeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAsp 460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3017 GCTAGC---TCTATCGCGACTTCACAAAGAGATGTTGCAAGAGATCATATGATGCTGAT 3073
QY 461 LysLeuProAspPheLysAsnValIleAspLysPheLysGlyAspThrLysLys 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3074 ATGGAAGTTGATGATTTGATGATTAAGATTAACATTAACAGAACTTTT---GACTACAAACT 3127
QY 481 TyrAlaAspPheValPheAspLysSerValIleProTyrSerAspLysPheHisAlaMet 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3128 GATATGACAAACAAGCTGGCTAATTAATGACACTACTATAGCATATAAACAAGCCTGGA 3187
QY 501 Leu-----LysSerMetAspLysGluLysPheAlaLys 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3188 ATACTATGCTGCTGCGATTTGAGAGCATCAAAAACCTTTGACCGCAGAAAGATGCTGAG 3247
QY 512 AlaIleGluLysAspProAlaValGluLeuSerLysSerValIleAlaIleAlaArgAla 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3248 GCCATTAGT-----GTTGCTGTGAGGGCC 3271
QY 532 IleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLysArgLeuPhePhe 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3272 TTGAGAGAGAGGACAAAG----- 3289
QY 552 AlaGlyLeuArgGluMetCysTyrProGlyArgAlaLeuProSerAspAlaAsnPheThrMet 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3290 -----GCCGTGTTCAAGAG 3304
QY 572 ArgMetSerTyrGlySerIleLysGlyGluProGlnAspGlyValArgTyrAsn--- 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3305 CGT-----AATGATATGATGACATGTTATACAAAGCCTTCGGCTTGATCCACGTT 3355
QY 591 ---TyrHisThrThrGlyLysGlyVal-----LeuGlnLysGlnAspProLys 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3356 ACATATCATCTCAATATTTGGGTTGCTACAAATCAAGGGTTGAAAAGACCTCATTTCAAT 3415
QY 606 SerAspGluPheAlaValGlnGluAsnIleLeuAsp-LeuPheArgThrLysAsnTyrGlu 625
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3416 ACCTTCCCTGGTGTGTTATGACCAGCTAATCCAGATTAAAGAGACAAACACAGTA-- 3473
QY 625 ValArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSer 639
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3474 -----ACAGGCCAGTTCTCAACTTGTCACTCTCA 3503

```

RESULT 12
 US-09-514-302-1
 ; Sequence 1, Application us/09514302
 ; Patent No. 638959
 ; GENERAL INFORMATION:
 ; APPLICANT: HATADA, Yuji
 ; APPLICANT: IGARASHI, Kazuaki
 ; APPLICANT: OZAKI, Katsuya
 ; APPLICANT: KAWAI, Shuji
 ; APPLICANT: ARA, Katsutoshi
 ; APPLICANT: ITO, Susumu
 ; TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PHOSPHATASE AND
 ; FILE REFERENCE: 2173-105P
 ; CURRENT APPLICATION NUMBER: US/09/514,302
 ; CURRENT FILING DATE: 2000-02-28
 ; EARLIER APPLICATION NUMBER: 08/952,084


```

Db 1121 ----- 1121
QY 455 gatgllleprolaasplysleuproaspilpehelsasnvallleasplyslsphy 475
Db 1122 -----GC 1123
QY 475 sglyasprhlylslytyrallaasphvealpheasplysSerValProTyrSerAs 495
Db 1124 GGGATACACCAACCAACCGGACCTACCTACGACGACGATGTGATCAGTTCCCGA 1183
QY 495 plysphehisalmetleuysSermetasplysglylsphalaallegluly 515
Db 1184 C-----AACCAAGACACCGCGGCTCGG--ACGCTCAACAG 1219
QY 515 sasproalavalgluleuserlySerVallealalaalarglaalleglinalas 535
Db 1220 CGATCCGGCG-----GCCCTCACCGGCGGCTC----- 1247
QY 535 palametalaasnlatyrallaile--glulysglylsargleuphealaglyle 554
Db 1248 -----GCCCTCCCTGCTCACACCGCGGCTACCGCTGCTTCTACGCG-- 1292
QY 554 uatgglumetyrProglyrargalaleuproSeraspalaasnphethrmetargmetse 574
Db 1293 -ACCGAGCAGTAC-----CTGCACAAAGCAGAC--GG 1321
QY 574 rtyrglySerlleysgllytyrgluProglinaasglyAlaTrpTyrAsn 590
Db 1322 TGAGGCGACGACAAAGGCAAGGCAAGACCG-----TACAAAC 1355

RESULT 14
US-09-008-172-1
; Sequence 1, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4310
; TYPE: DNA
; ORGANISM: Streptococcus mutans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5)...(4295)
; US-09-008-172-1

Alignment Scores:
Pred. No.: 0.736 Length: 4310
Score: 102.00 Matches: 154
Percent Similarity: 34.30% Conservative: 129
Best Local Similarity: 18.67% Mismatches: 280
Query Match: 2.74% Indels: 263
DB: 3 Gaps: 41

US-10-008-355-2 (1-712) x US-09-008-172-1 (1-4310)
QY 20 glyValalalaasplyslymetrpleuasnngluenuasn----- 35
Db 1463 GGTGTTGATTAAGAGTGGAAAAAT-----GCCATTAAATCATCTTTTCATTGAAGCT 1516
QY 36 -----Gingluasnleuasnargmetarggluleuglypethrleu 49
Db 1517 TGGTCAGATAATGATCCCAATCAATTAAGATACCTAAG-----GGTGCAATTA 1567

```

```

QY 50 ProteuaspSer-----LeuTyrSerPheasplysPro----- 60
Db 1568 CCAGTTGATTAATAACTGCCCTATCGCTTTATATGCGCTTGACGCGCTCTTGAAAAA 1627
QY 61 SerllealasnlaValallephegllygllycysthrlylethrValSerAsp 80
Db 1628 GATGCAAGCAATTAATAATTCGACGCGACTTAGCCTGTCAATACAAATAGC--- 1684
QY 81 glnglyleuilepethrAsnHisScysglytyrlylaalleuSerInserThr 100
Db 1685 -----TGATTAACCGTTACGCGAAGGTAA--AATAGGAACGTATG 1726
QY 101 Valasphisaplyrleuarg-----AspGlyPheValSerarghrMetglyglu 118
Db 1727 GCTAATCATATTTTAAACCGCTCACGACGACGATGCAACGATG----- 1774
QY 119 leuproilleproglyleuserVallytyrleuarglylsvallyThrAsplys 138
Db 1775 -----ATTGCTAAATTTTAAAGCT-----CAG 1798
QY 139 ValgluagluleuylslylethrAspglumetgluargleuarglylainglu 158
Db 1799 ATTAATCCCAAAACAGATGTTGACCTTACTTTGATGAATGAAGCAACGCTTAAG 1858
QY 159 ValCysglnglu-----leuAlaLysLysgluasnlaaspgluasnnglu 175
Db 1859 ATCTACAAATAGACATGCGTCAAGCTAAGAAAAGATCACACAAATCAATATTCGACA 1918
QY 176 lleValgluProphetyrSerAsnsglytyrPheleuileValtyr-----AspVal 193
Db 1919 GCCATGCTTTGATGCTGCTCAATTAAGATTCTATACAGTCTTTATTTGATGATATG 1978
QY 194 PheLysaspValargmetvalPheAlaProSer----- 205
Db 1979 TACAGTATGACGCTCAATATATGCGATAAATCCCTTATATGATGATGATATG 2038
QY 206 -----SerValglyLysPhe 210
Db 2039 TTATTAAGCAGCATTAATATATGCCGCGGTGCTCAAGACATGAAGATCACTATGTT 2098
QY 211 GlylyaspThrAspAsntrpmetrpproArgHisThrlyaspPheSerValPhearg 230
Db 2099 GAAGGTATTAAGATCATATGATGGAT--TATACAGCGTTTGTACCTTGCTTGT 2155
QY 231 ValtyrAlaaglalaasp----- 236
Db 2156 TATGTTACAGAGCTAATGAGCTACAGATCAAGCAAGCAAGCAACTAAACAAAGCA 2215
QY 237 -----AsnargProAla----- 240
Db 2216 ATGGCTGTCAATTACACGACAAATACCTAGCCTTAATTAATGAATCAAAATGATTAAT 2275
QY 241 -----gluTyrSerlyAspAsnlyspProTyrLysProValtyrPheAlaAlaVal 257
Db 2276 GTCAATATGGGGGTGGGCAATTAATAATCAAGAGTACCGTCCGCTTAACAACTAA 2335
QY 258 ---SermetcnglytyrlysaalaspAspTyrAlamethrllleuglyPheproglySer 276
Db 2336 GATGTTGATCAAGCTACATCTTGATGCGCGCTCAATACCTT--TATCGCAAAAGC 2392
QY 277 ThrAsparglyrleuThrserTrpGlyValgluAsparglylleuasnngluasnPro 296
Db 2393 AATGATTAAGAGCAATTAATGCTTTGATGCTAGTGC--ATTCAAAGTTACTCGAAATCG 2449
QY 297 ArgllleuValargglylleysglnglyleryrplysgluAlametSerAlaAsp--- 315
Db 2450 -----CAAGTATACAGTTAT--TATGCGGTTTGGTTCCAGTAGAGAGCTAGATTAAT 2500
QY 316 GluAlaThrarglyleystyrAlaSerlystyrAlaInserAlaasnTyrTrpLysAsn 335
Db 2501 CAAGATGTTGCTGTAGCAGCAAGCAATTAAGCAATCTACTGCTCAAGCTACAGATCA 2560
QY 336 Serlleuglymetasn-----ArgglyleuAlaargleu----- 346

```

```

Db 2561 TCAGTCTCTGATTCGAATTTGATTTAGCAAGGTTTCTCAACTTCACAGATTTTGTGA 2620
OY 347 -----AspValIleGlyArgLysArgAla----- 354
Db 2621 ACGAAGATTCAGACTATCTAATAAGAGATTGCTCAAAATGCTCAACTCTTCAAAATCT 2680
OY 355 -----GluGluArgAlaPhe 359
Db 2681 TGGGGTGTCTCTTTCGAATGGACCGCAATATGTCTCTTCGAAGATGTTCTTTT 2740
OY 360 AlaAspTrpIleArgLysAsnGly----- 367
Db 2741 CTAGACTCTATTATTCAAATATGTTATGCTTTGAGAGATGTTATGATCTTCTATGACT 2800
OY 368 LysSerAlaValIleGly-----AspValIleuSerSerLeuLysAlaTyrLys 384
Db 2801 AAGATTAACAATACGATTCCTCAGCAAGACATGATTAAAGCGTTAAACCTCTGCATATA 2860
OY 385 GluGlyAlaLysAlaAsnArgLysLumethTrpTyrLeuSerGluThrLeuPheGly----- 402
Db 2861 ACGGTAATCAGGTTATTGCGGAT-----TGGGTACCAAGATCAATCATATATCTCCG 2914
OY 403 GlyThrGluValIleArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAla 422
Db 2915 GGCAAGAAGTGTGA-----ACGGCTACACGCTGTCAACGAT 2950
OY 423 His-AlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAs 442
Db 2951 TATGGTAGATTCGCAAGACCTCTGAATCAAAATACCTTATGCTGCCACA----- 3005
OY 442 PArgLysValLeuProAlaMetLeuAspIleValArgArgGlyIleProAlaAspLysLe 462
Db 3006 -----CTAAGAGTAATGTAAGTATCAAGCAAGTATG----- 3041
OY 462 uProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspTrpLysLysTyrAl 482
Db 3041 ----- 3041
OY 482 aaSPheValPheAspLysSerValIleProTyrSerAspLysPheHisAlaMetLeuLys 502
Db 3042 -----GCGGTCTTCCCTTAGTAAGTACGCGCTGCTAAGTACCTTACT 3082
OY 502 sSerMetAspLysGluLysPheAlaLysAlaIleGluLysAspProAlaValGluLeuSe 522
Db 3083 ATCTTTAACCGCACCAATTTCAATGTGTAG-AAGATTGATCCAAAGCAAGAAAA----- 3136
OY 522 rLysSerValIleAlaAlaAlaArgAla----- 531
Db 3137 -----ATCACAGCATGGAGCAAAATACTTCANTGGGACAAATATTCTAGGCGG 3186
OY 532 -----IleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGlyLys 547
Db 3187 TGGTGTGGTATTATGTTTAAAGATAATGCTAGTGTATTAATCTTTGAAGCTAAAGCGGA 3246
OY 547 sArgLeuPhePheAlaGlyLeuArgLysLumethTyrProGlyArgAlaLeuProSerAspAl 567
Db 3247 TCACAACTATCTGCCA-----AAACAGATGACTAACAAAGAAAGCTTCACGTGTTTGT 3300
OY 567 aaSnPheTrpMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAsp----- 585
Db 3301 TATATGATGGCAATGGAGATGCTTTCATCTAAGTGTGTTCAAGCAAGCAAGCGCTT 3360
OY 586 -----GlyAlaTrpLysAsnTyrHisThrThrGlyLysGlyValLeu-- 599
Db 3361 TGTTCAGATGCCCAAGAAAGTGGTATTACTTGTATATATATGGCCATATGCTTTATGCG 3420
OY 600 -----GluLysGlnAspProLysSerAspGluPheAlaValGlnGly 613
Db 3421 CTTCACAGACGATAAATGGGAGAGTCAATCTTTTATCAATATGGTGTTCATATGCGTGA 3480
OY 613 uAsnIleLeuAsp-----LeuPheArgThrLysAsnTyrGlyArgTyrAlaGlnLysGlnGly 632

```

```

Db 3481 ATCTTCTTGGAAAAACGCTGATGGACAGCAAGACTATTTGCTCATCTAGAAATAGATA 3540
OY 632 nLeuHisIleAlaPheLeuSerAsnAspIleThrGlyLysAsnSerGlySerProVa 652
Db 3541 TAGTATATGTTTATTTATCTTTGATATATGAT-----AGTAATGGCGTTA 3585
OY 652 lPheAspLysAsnGlyArgLeu--IleGlyLeu--AlaPheAspGlyAsnTrpGlnAl 670
Db 3586 TTTTATGCCAGTGGATGATGCTGTAGCTTTGAAACAAATTAACGCCAATACCCAGTA 3645
OY 670 aMetSerGlyAsp 674
Db 3646 CTTTGATCAAGAT 3658

RESULT 15
US-09-210-361-5
; Sequence 5, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 4310
; TYPE: DNA
; ORGANISM: streptococcus mutans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5)....(4295)
US-09-210-361-5

Alignment Scores:
Pred. No.: 0.736 Length: 4310
Score: 102.00 Matches: 154
Percent Similarity: 34.30% Conservative: 129
Best Local Similarity: 18.67% Mismatches: 280
Query Match: 2.74% Indels: 263
DB: Gaps: 41

US-10-008-355-2 (1-712) x US-09-210-361-5 (1-4310)
OY 20 GlyValAlaLysAlaAspLysGlyMetTrpLeuAsnGluLeuAsn----- 35
Db 1463 GGTGTGATTAAGATGAGAAAAT-----GCGATTATATCATCTTCCATTTTACAGCT 1516
OY 36 -----GlnGluAsnLeuAspArgMetArgLysLumethLysPheThrLeu 49
Db 1517 TGGTCAGATTAATGATCCCAATACATATAAGATTAAG-----GTCACAAATTA 1567
OY 50 ProLeuAspSer-----LeuTyrSerPheAspLysPro----- 60
Db 1568 CCGATTGATTAATAAATCGCGCTATCGCTTTATATGCTTTGACCGCTCTTGAAAAA 1627
OY 61 SerIleAlaAsnAlaValIlePheGlyGlyGlyCysThiGlyIleThrValSerAsp 80
Db 1628 GATGCAAGCAATAAATGAATTCGACGCGGACTTGAACCTGTGATACAAATAGC--- 1684

```


QY 81 GlnGlyLeuIlePheThrAsnHisCysGlyTyrGlyAlaIleGlnSerGlnSerThr 100
Db 1685 -----TTGAATTAACCGTTTCAGCTGAAGGTAA--AAATGTAAGCATG 1726
QY 101 ValAspHisAspTyrLeuArg-----AspGlyPheValSerArgThrMetGlyGlu 118
Db 1727 GCATACATATTTTATATCCCGCTCAGACAGTCAAGTCCAAACGGT----- 1774
QY 119 LeuProIleProGlyLeuSerValTyrLeuArgLysIleValLysValThrAspLys 138
Db 1775 -----ATTGCTAAATAATTATTAAGCT-----CAG 1798
QY 139 ValGlnGlyGlnLeuLysGlyIleThrAspGlnMetGlnArgLeuArgLysAlaGln 158
Db 1799 ATTTAATCCCAAAACAGATGGTTGACCTTACTTGGATGAATGAAGCAAGCTTTAAG 1858
QY 159 ValCysGlnGlu-----LeuAlaLysGlnGlnAsnAlaAspGlnGlnLeuCys 175
Db 1859 ATCTACAAATGAAGCATGCGTCAGCATAGAAAAAGTACACACATCCATATTCGACA 1918
QY 176 IleValGlnProPheTyrSerAsnAsnGlnTyrPheLeuIleValTyr-----AspVal 193
Db 1919 GCCTATGCTTGGATGCTGCCAATAAAGATTCTATTACAGCTCTTATTATGCTGATATG 1978
QY 194 PheLysAspValArgMetValPheAlaProPheSer----- 205
Db 1979 TACAGTATGACGCTCATATATGCGCATAAATCCCTTATTATGATGCTATTGATCT 2038
QY 206 -----SerValGlyLysPhe 210
Db 2039 TTATTAAAGCAGCTATTATATATGCGCGGTGCTCAGACATGAAAGATCACTATGTT 2098
QY 211 GlyGlyAspHisAspAsnTyrPheTyrProArgHisThrGlyAspPheSerValPheArg 230
Db 2099 GAAGGTATATAAAGCTATATGATGGATGAT--TATACAGCGCTTTGACTTCTGTTGCT 2155
QY 231 ValTyrAlaGlyAlaAsp----- 236
Db 2156 TATGGTACAGAGCTATATGAACTACAGATCAAGGCACTGAAGCACTAAACACAAAGA 2215
QY 237 -----AsnArgProAla----- 240
Db 2216 ATGGCTGTCAATTACACCAATTAACCTTAGCTTAATTGAATCAATAATGATAATATT 2275
QY 241 -----GlnTyrSerLysAspAsnLysProTyrLysProValTyrPheAlaIleVal 257
Db 2276 GTCAATATGGGGGCTGCGCATATAAATCAAGAGTACCGCTCCTTAACAACATAAA 2335
QY 258 ---SerMetGlnGlyTyrLysAlaAspAspTyrTAlaMetThrIleGlyPheProGlySer 276
Db 2336 GATGGTTTGACAAAGCTACCTCTGATGCGCTGCTAAATCCCTT--TATCGCAAAAGC 2392
QY 277 ThrAspArgTyrLeuThrIleSerTyrGlyValGlnAspArgIleGlnAsnGlnLysAsnPro 296
Db 2393 AATGATTAAGGAGAAATAGTCTTGTAGCTAGTGAC---ATTCAAGGTACTCTCAATCCG 2449
QY 297 ArgIleGlnValArgGlyIleLysGlnGlyIleTyrLysGlnAlaMetSerAlaAsp-- 315
Db 2450 -----CAAGTATCAGGTAT--TTAGCCGTTGGGTCCAGTAGAGGCTAGATGAAT 2500
QY 316 GlnAlaIleThrArgIleLysTyrAlaSerLysTyrTAlaGlnSerAlaAsnTyrTyrLysAsn 335
Db 2501 CAAAGATTTGCTGTAGACAGCAAGCAATAGGCAAAATGCTATGGTCAAGTCTACGAATCA 2550
QY 336 SerIleGlyMetAsn-----ArgGlyLeuAlaArgLeu----- 346
Db 2561 TCAAGTGTCTGTGATTTGCAATTGATTACGAAGGTTTCTCAAACTTCCAAAGATTGTGA 2620
QY 347 -----AspValIleGlyArgLysAlaIle----- 354
Db 2621 ACGAAGATTCAGACTATTAATAAGAAAGATTGCTCAAAATGTCCAATCTTCAAAATCT 2680

QY 355 -----GlnGluArgAlaPhe 359
Db 2681 TGGGGTGTCACTTCTTGAATGCGACCCGCAATATGCTCTTCTGAAGATGTTCTTTT 2740
QY 360 AlaAspTyrIleArgLysAsnGly----- 367
Db 2741 CTAGACCTTATTTTCAAAATGGTATGCTTGAAGATGGTTATGATCTTGGTATGAGT 2800
QY 368 LysSerAlaValTyrGly-----AspValLeuSerSerLeuGlnLysAlaTyrLys 384
Db 2801 AAGATTAACAATTCGGTCTTCAGCAAGACATGATTATATCAGTTAAAGCTCTGCATAAA 2860
QY 385 GlnGlyAlaLysAlaAsnArgGlnMetThrTyrLeuSerGlnThrLeuPheGly----- 402
Db 2861 ACGGTTATCAGGTTATTCGGGAT-----TGGGTACCAAGTCAAACTTAATCTTCCG 2914
QY 403 GlyTyrGlnValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAsnProAspAla 422
Db 2915 GGCAAAGAAGTCGTA-----ACGGCTACACGTGTCACAGAT 2950
QY 423 His-AlaGlyIleLeuLysSerLeuAspLysTyrLysAspTyrLeuProSerLeuAs 442
Db 2951 TATGGTAGATATCGCAAAAGCTCGAATCAAAATACACTATGCTGCCAACACA----- 3005
QY 442 PArgLysValLeuProAlaMetLeuAspIleValArgArgIleProAlaAspLysLe 462
Db 3006 -----CTAAGACTAATGGTAAGATTATCAACGAGATATG----- 3041
QY 462 uProAspIlePheLysAsnValIleAspLysLysPheLysGlyAspHisLysTyrAl 482
Db 3041 ----- 3041
QY 482 AspPheValPheAspLysSerValValProTyrSerAspLysPheHisAlaMetLeuLys 502
Db 3042 -----CGCGTCTTCTTATGATGACCTGCTGATACCTAGT 3082
QY 502 sSerMetAspLysGlnLysPheAlaLysAlaIleGlnLysAspProAlaValGlnLeuSe 522
Db 3083 ATCTTTAACCGCAGCGCAAAATTTCAAAATGATAG- AAGATTGATTCAAAGCAAAAA----- 3136
QY 522 rLysSerValIleAlaAlaAlaArgAla----- 531
Db 3137 -----ATCACAGCATAGGAAGCAAAATACTCAATGGGCAAAATATTGAGCGCG 3186
QY 532 -----IleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGlnLysGlyLys 547
Db 3187 TGGTGTGTTATTCCTTAAAGATATGCTAGGATAAATACTTTGAACGTGAAGAGGGA 3246
QY 547 sArgLeuPhePheAlaGlyLeuArgGlnMetTyrProGlyValArgAlaLeuProSerAspAl 567
Db 3247 TCAAACCTATTCGCA-----AAACAGATGACTAACAAAGAACCTTCGACTGGTTTGT 3300
QY 567 AsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlnProGlnAsp----- 585
Db 3301 TAAATGATGGCAATGAGTACTTCTATTCACATAGAGGTTATATCAACCAAGAACACTT 3360
QY 586 -----GlyAlaTyrTyrAsnTyrHisThrThrGlnLysGlyValLeu-- 599
Db 3361 TGTTCAAGATGCCAAAGAAATGCTGATTTACTTGAATAAATGAATGACCATATGCTTATG 3420
QY 600 -----GlnLysGlnAspProLysSerAspGlnIlePheAlaValGlnI 613
Db 3421 CTTAACAGACGTAATAGCGCAAGTGCATTAATCTTTTATCAAAATGGTCTCAATTTGCGTGA 3480
QY 613 uAsnIleLeuAsp---LeuPheArgTyrLysAsnTyrGlyArgTyrAlaGlnLysGlnI 632
Db 3481 ATCTTCTTGGAAACCGCTGATGGCAGCAGAACTATATTTGGTATCTAGGAATATAGATA 3540
QY 632 pLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlnLysSerGlySerProVa 652
Db 3541 TAGTAATGGTATATATCATATTGATTAATGAT-----AGTAAGTGGCGCTTA 3585
QY 652 lPheAspLysAsnGlnLysArgLeu---IleGlyLeu---AlaPheAspGlnLysnTyrGlnAl 670

Db	3586	TTTTGATGTC	CGTGGATCAT	GGCTGTAGGTTT	GAACAAATTA	ACGGCAATAC	CGCAGTA	3645
			:::	:::	:::	:::	:::	
Qy	670	ametserclyasp	674					
Db	3646	CTTGATTCAC	AGAT	3658				

Search completed: October 18, 2002, 01:09:38
Job time : 4113 secs


```

Db 175 FSFAGNTYRVFKNLEIKDVLRYAPQGSVCKFGGDVNMMPRHGTDFSEYRAVVGKDG 234
Oy 238 RPAEYKONKRYKPYFAAAMOGKADDDYAMTIGFPGSTDRYLSMVEEDRIENENPR 297
Db 235 KPASCKENIPYRPHMKESDQPLGDDFVWAGYFEGRTNRYALV-----AEENIAH 288
Oy 298 IEVNGIKO-----GIMWEASADOATRIKASKYQAOSANWKNISGMNGLARLDVIGR 351
Db 289 WTYPIQGFKNLIALIEAASKONPDIOVKYASTLAGINNTSKNFDGQDDEFRRLNALGQ 348
Oy 352 KRAEBRPAADIRKNG-----KSAYGDVLSLEKAYKEGAKANREMYLSETEFGTEV 407
Db 349 KQSEETAVLAWLKQOQIGRHEALAHQTLVDLTDY-----KANODRDFVLGO-FENGSGVI 403
Oy 408 RPAQFANLA---TNPDAH-----AGILKSLDKRYKDYLSLRKY-----445
Db 404 GVAANLYRLAERTKSDQRENGIOERDLPTIEGNLKOME---RRYLEDKMRQMOQWLT 460
Oy 446 ---LPAMLDIVRRIRPA-----DKLPDIFKNVIDKKFGDTRKYADFVDEKSVYPSD 495
Db 461 EYNKLP-----VKQRYAIDVWLDGIPATLRLGDTKLSSSEER-----500
Oy 496 KFHAMLSKDEKFAKAIKQPAVELSKSVIAARAIOADMANAYAEKGRLEFAGLR 555
Db 501 ---LKFENADRAAFESSODPAIRYAVAIIMPALLEIERONKIRTEGELKARPYTLQALA 555
Oy 556 EMYP---GRALPSDANFTMRSGISKEYEPODGAANYHTTGKGLVLEKODKSPDEFAQOE 613
Db 556 DYNSHGFEVYPPDANSSIRITFGHAKGSPKDYETPTFLLQGYMAK-NGVEFEDSPK 614
Oy 614 NILDFTKNGRYAEV---GOLHIAFLSNNDITGNGSGSPYEDKNGRLIGLAFDGNWEAM 671
Db 615 SLINAKKASVIANLADQIGIVPVPVFLSDLDITGNGSGSPYADAHGKLVGLAFDGNMESV 674
Oy 672 SGDIEFEPDLORTISVDIRYVLEMDKMGQCPRLIOEKL 711
Db 675 SSNWVEPVMTRTIAVDSRYVQWIMTEVAPAPHLKELENL 714

```

RESULT 2

A60999
 alpha-amylose (EC 3.2.1.1) precursor - Micrococcus sp. (strain 207)
 C:Species: Micrococcus sp.
 C:Date: 31-Dec-1993 #sequence.revision 31-Dec-1993 #text.change 15-Oct-1999
 C:Accession: A60999
 R:Kimura, T.; Horikoshi, K.
 FEMS Microbiol. Lett. 71, 35-42, 1990
 A:Title: The nucleotide sequence of an alpha-amylose gene from an alkalophilic thermophilic
 A:Reference number: A60999
 A:Accession: A60999
 A:Molecule type: DNA
 A:Residues: 1-1104 <KIM>
 A:Cross-references: GB:X55799; NID:g296762; PIDN:CAA39321.1; PID:g296763
 C:Function:
 A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
 A:Pathway: glycogen/starch degradation
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation
 F:1.32/Domain: signal sequence #status predicted <SIG>
 F:3.3-1104/Product: alpha-amylose #status predicted <MAT>

Query Match 3.6%; Score 133.5; DB 2; Length 1104;
 Best Local Similarity 20.8%; Pred. No. 1.8;
 Matches 138; Conservative 94; Mismatches 259; Indels 173; Gaps 35;

```

Oy 141 GOLGIDTEMERLKAQVCEQLAKKENADENQLCIVPEFISNNXYFLIVDVEDVAMV 200
Db 384 GDFGIGTIDKLYL-----DELGVNTIWI-SPVVENIKY-----DVRYX 420
Oy 201 FAPSSVCKFGGDVNM---MMPRHGTGDSVFR-VYAGADNPAEYKONKPYFAAV 257
Db 421 ETSPEYIYRHYWANNFGLNP-HGTMEEFHDLIDGADHDKMKIMVD-----VYV 470
Oy 258 SMOGY-KADYDAMT---IGFPGSTDRYLS-----MGVEDRIENEN 294

```

```

Db 471 NHTGYGLKEIDGSVTPNPAGYPSDADRARFSDLLRGADVGTDEYVGLACLPPDITEEDP 530
Oy 295 NPRIEVRKIGKIGKKEAMSAOATRIKY-----ASKYASANWKNISGMNGLARLDVIG 350
Db 531 NVRKQIID-WQTDWIEKATYENGNTIDFVRVDTYKAHVEDAWMOFKNALTEKMEPFKMG 589
Oy 351 KRAEBRPAADIRKNGSAVY---GDVLSLEKAYKEGAK-----ANREMYLSE 398
Db 590 EA-----WGAKVDNTLGILETGYMDSLDDGCFKETAHPNGSLAANASITARNA 640
Oy 399 TLFGTEYVR-----AQFANALATN-----PDHAGILKSLDKRYK 435
Db 641 KLDWTATLQGLFSGHDEGFLSHLAGDKQLQVATLQATKAGQPVLYYG--EELGOTGA 698
Oy 436 DYLSLDRKVLPAHLDIVRRIRPADKLPDINKNYIDKKFKDTRKYADFVDEKSVYPSD 495
Db 699 NNYPQYDNR---YDFAMDQVEGENEILAHYTKILNR-EGYSKYFAK--GERTLVGSD 750
Oy 496 KFHAMLSKDEKFAKAIKQPAVELSKSV---IAARAIOADMANA-YAIEKGRLEF 551
Db 751 KDQFLPSRDYQDQKYYVGLVAEE-SKAVTLTVDSADAVTDAVSGTEYRATAGK-----805
Oy 552 AGLREMPGRALPSDANFTMRSGISKG-----YEPDGAANY 590
Db 806 ---VNLTLPKADPGTVLLT--VEGGINITGVAKDNGEVVVELVPENNIRIHYKREDNMYKN 861
Oy 591 YHTGKGLVLEKODKSPDEFAVOENILDFRTKNGRYAE-----NGOLHIAFLSNNDITG 646
Db 862 Y---GAMLMNDVASPSANMPGATMFE--KTDSTGATIDVPLADGAKNIGLV-MDTIAG 915
Oy 647 NSGSPVEDKNGRLIGLAFDGNWEAMSGD--TEFEP-DLORTISVDIRYVLEMD-----697
Db 916 DAGKDGDKGFTTSSPQANEIWKOGSDKYTYTEPVLDLPAN-TYRIHYTRBAVYDDPCI 974
Oy 698 -KMG 700
Db 975 WNWG 978

```

RESULT 3

D97933
 valine-tRNA ligase (EC 6.1.1.9) [imported] - Streptococcus pneumoniae (strain R6)
 C:Species: Streptococcus pneumoniae
 C:Date: 22-Oct-2001 #sequence.revision 22-Oct-2001 #text.change 02-Nov-2001
 C:Accession: D97933
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.;
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
 A:Reference number: A97872; MUID:21429245; PMID:11544234
 A:Accession: D97933
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-883 <KUR>
 A:Cross-references: GB:AE07317; PIDN:AAK9296.1; PID:g15458063; GSPDB:GN00174
 C:Genetics:
 A:Gene: valS
 C:Superfamily: valine--tRNA ligase
 C:Keywords: ligase

Query Match 3.6%; Score 133; DB 2; Length 883;
 Best Local Similarity 20.4%; Pred. No. 1.4;
 Matches 104; Conservative 65; Mismatches 166; Indels 176; Gaps 24;

```

Oy 192 DVF-----KDYRMVFAPPSSVCKFGGDVNMMPRHGTGDSVFRVYAGADNRPAPYS 243
Db 26 DVFRPSDQKAKKPYSIAPPNVTGKL-----HLG-----HA 57
Oy 244 KDNPKYKPYFAAVSMOYKADYDAMTIGFPGSTDRYL-TSWGVEDRIENENPRIEYRG 302

```

Db 58 WDT-TLDDIIIRKMOGFD-----TLMLPGMDHAGIATQAKVERLRGEGISRYDLGR 110
 QY 303 -----IKGIMKEAMSAOQATRIKYA-SKYAOSANYWKNISGMNGLARLDYIGRKAREER 357
 Db 111 ESFTLKWEMKD-----EYATTIKQMGKMGISVDYSREPTLDEGLS-----KAVRK 158
 QY 358 AFADMIK-----NGKSAVYGD-----VLSSLEKAYK-----EGAKANEMT 394
 Db 159 VFVDLYKKMGIYRGEFIINNDPAPARTALSDIEVYHKDEGAFYHMNMYMLEGSALEVAT 218
 QY 395 YLSTLEFGTEGVVRFQAFANALATNPDAHGIKSLDDKKDYLPs-----LDRKVLPAW 449
 Db 219 TRBETMGDV-----AIAVNPE-----DPRYKDLGKXVILPIANKLIPIY 259
 QY 450 LD-----IYRRRIIPADKLPDIFKNVIDKKFKDDTKKYPADVPVDPKSVYPSDK 496
 Db 260 GDEHADPEFTGYVKTIPADHPDNLVQGRHNPQVAINMDDGTMLNVEFSGM-----DR 316
 QY 497 FHMLKSMDEKEFAKIEKPAVELSKSVIAARA-----IQAD 535
 Db 317 FEAR-----KAVAKLEIGALVYIEKRVHSVGHSERTGVVPERLSTOMFVKMDOLAKN 371
 QY 536 AMANAVYIEKGRLEFAGLRPMTPGRALPS-----DANFTMRSSYSGINKYEPQDG 586
 Db 372 AIANQDTEDK-----VEFYPRFNDFLQWENYHDMVISROLMWGH-----QIP 416
 QY 587 AMYVHTTGKGVLEKQDPKSDPEFAVOENIID 617
 Db 417 AMYN--ADGEMVGEAPESGDGTQDDEDVID 445

RESULT 4

A36866
 microbial collagenase (EC 3.4.24.3) precursor - Clostridium perfringens
 C:Species: Clostridium perfringens
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C/Accession: A36866
 R:Matsumita, O.; Yoshihara, K.; Katayama, S.I.; Minami, J.; Okabe, A.
 J. Bacteriol. 176, 149-156, 1994
 A:Title: Purification and characterization of a Clostridium perfringens 120-kilodalton
 A:Reference number: A36866; MUID:94110220
 A:Accession: A36866
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1104 <MAT>
 A:Cross-references: GB:D13791; NID:9440850; PIDN:BA02941.1; PID:9440851
 C:Genetics:
 A:Gene: COLA
 C:Superfamily: microbial collagenase
 C:Keywords: hydrolase; metalloproteinase

Query Match 3.58; Score 129; DB 1; Length 1104;
 Best Local Similarity 18.2%; Pred. No. 3.6;
 Matches 142; Conservative 106; Mismatches 276; Indels 258; Gaps 36;
 QY 29 WLNLNLEQENL-----DRMRELFTPLDLSYSPDKPSIANAVYIEGGGC-----GIYVSD 80
 Db 419 WAKSEVKAQFMRYVQNDKALEGNPDILTVIYNSPEETKLNRINFGSTDNCGIYEN 478
 QY 81 QGLFTNHHCYGAIGS-QSTVDDH-----YLKDFVSRMTGELPLPGL-----SVYELRKIV 133
 Db 479 IGFETTERPEESIIYLEELFRHEFTHYLGGRV-----VPGMMGGEGEYQGSVL 529
 QY 134 KYIDKVGQLKGTIDEMERLKAQVCOELAKKENADENQLCIYEPYSNNEYFLIYDV 193
 Db 530 TWVEEGTAIEFFAGSTRDGIRKRSYVQGLAYDRNNRMSLYGLHARYGSDPF-----583
 QY 194 FKIDRMVFAFPSSVKGKGTIDNMMPRHGDSFVFRVYAGADRPAEYSDKNPKPYV 253
 Db 584 -----NKGFLSNMYNNKGMF-----NKMNYIKNN-----611
 QY 254 PAAVSMOGYK-----ADYAMTIGFPGSTDRYLTLSGVEDRIENENNPRIE--VGIGK 304

Db 612 ----DVSGYKDYIASMSDSYGLNDKYODYMDSL-----NNIDNLDPVLSDEVYNG-- 659
 QY 305 QGIMKEAMSAOQATR-IKYASKYAQ-SANYWKNISGMNGLARLDYIGRKAREERAFADW 362
 Db 660 ----HEAKDINELITNDIKESVNIKOLSSNVEKSGPFTTYDGRGYVGGRSGEEN--DW 712
 QY 363 IRKNGKAVYGDVLSLEK-----AYKEGAKANREMYTSEITLFGTEGVVRFQAF 412
 Db 713 KDMASK---LNDMLKELSKSMNGYKVTYAFVHKKYDENQNYVYDVFGHMT-----763
 QY 413 ANALATNPDAHA-----GIKSLDDKYKDYLPISDRKVLPAMLDIYRRRIIPADKLPDIFK 467
 Db 764 ----DTMTDVHVNKEPKAVIKS-----DSSV-----IYEEIN-----FD 794
 QY 468 NVIDKKRGDTKKY-----ADPEPKSVYVPSDKFHMLKSM-----K 506
 Db 795 GTESKDDDELKAYEMDFGDEKSNKAKHKKYK-----TGEYEVKLTVDNNGGINTG 849
 QY 507 EKFAKALEKDP-----AVELSKSVIAARAIOADAMANAIEKGR---548
 Db 850 SKKIKVYEDKFEVYINSEPNDFEKANQIAKSMMLYKCTISEDYSDKYTFDVAKKGNV 909
 QY 549 -----LFFAGLRMY-----PGRALPSDANFTMRSSYSGINKYEPQDG 586
 Db 910 KITLNLNSVGIPTWLYKEGDLNMYLYANGNDGTLEKGTLEPGRVYLSVYTYDNOG 969
 QY 587 AMYVHTTGKGVLEKQDPKSDPEFAVOENIIDLFTKRYNGRAENGQILIAFLSNNDITGC 646
 Db 970 A--YTVANVGNLKNKYEKEDAIKE-----VEN-----NND-----999
 QY 647 NSGSPVFDK-----NGRLIG-LAFDGNWMAMSGDIEPEPLQRTI-----SVDIRYVLFM 695
 Db 1000 -----FDKAMKVDNSKIVGTTLSNDLKIDYISIDIQPSDLNIVENLDNIKMMILYS 1053
 QY 696 ID 697
 Db 1054 AD 1055

RESULT 5

T33813
 hypothetical protein VC5.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T33813
 R:Tin-Wollam, A.M.; Wohldmann, P.
 submitted to the EMBL Data Library, November 1998
 A:Description: The sequence of C. elegans cosmid VC5.
 A:Reference number: Z21417
 A:Accession: T33813
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1440 <TIN>
 A:Cross-references: EMBL:AF106581; PIDN:AACT8208.1; GSPDB:GN00023; CESP:VC5.3
 A:Experimental source: strain Bristol N2; clone VC5
 C:Genetics:
 A:Gene: CESP:VC5.3
 A:Map position: 5
 A:Introns: 15/3; 27/2; 258/3; 382/3; 506/3; 630/3; 754/3; 879/3; 1005/3; 1410/3

Query Match 3.48; Score 125; DB 2; Length 1440;
 Best Local Similarity 20.1%; Pred. No. 9.8;
 Matches 108; Conservative 72; Mismatches 162; Indels 194; Gaps 30;
 QY 132 IVK-VTDKVEGQLKGTIDEMERLKAQVCOELAKKENADENQLCIYEPYSNNEYFLIV 190
 Db 237 LVKGVVDROGEVYIAEKM-----LSVCGEYVKESTRKR-----271
 QY 191 YDVKDYRMVFAFPSSVKGKGTIDNMMPRHGDSFVFRVYAGADRPAEYSDKNPKY 250
 Db 272 ----RELEAFAQ-----DFVKWMTPEQLDITALLA-AGKDSVOAKVE-----311
 QY 251 PVYFAAVSMOGYKADYAMTIGFPGSTDRYLTLSGVEDRIENENNPRIEYVGIKGIWKE 310

```

Db 312 --FFGOLPADQOK---TLTEEFKCKCKYFTPLMTSELE-----KIKTLKGD--KE 356
Qy 311 AMSADATRIKAKSAKQASANYMKNISIGNRGLARLDYIGRKAERARADMTKRNKGA 370
Db 357 AAGA-----LVKG-----VVDREGKAVAAE----- 378
Qy 371 VYGDVLSLEKAYKEGAKANREMTYLSFTLFGTEVYRFAQFA-----NALATN 419
Db 379 ---KMLTVCGEVYKKDSSRRRRELE-----AEQDFYKMMTPPOLGINDLTKT- 422
Qy 420 PDHAGILKSLDDKYDY---LPSLDRKVL-----PAN-----LDIVRRIPA 459
Db 423 ---AGKDESIDDKVEYFALLPSPDOQKTLTEEFKCKCKYFTPLMTSELEKTI-KTMD 477
Qy 460 DKLP--DIRKNYIDKKFKGDTKRYADVFVDKSVVPYSDK-----FHMLKSMDE 507
Db 478 DKVAGALVKGVYDRO--EEVKAKAIKMLSVCGEYVKDKRRKRELEAFAKDFVHMTPPE 536
Qy 508 KFA-----KAIEKDPAVELS-KSVIAAARAIQADAMANAAYALEKGR-RLFFAGLREMYPG 560
Db 537 QLGDITALKAGKDSVQAKKVEFFGQLPADQOKTLTEEF---KCKCKYFTPL----- 587
Qy 561 RALPSPANTMMSYSGISIKYEPQDQAMYNHTGKGVLEKODPKSDEFRAVOENIL 616
Db 588 --MTSEDLAKIE---TLKGDKKAAAGALV-----KGVVDROE--GDVAKAIKMKL 629

```

RESULT 6

P98228

1,4-alpha-glucan branching enzyme (glucogen branching enzyme) [imported] - Agrobacterium

C:Species: Agrobacterium tumefaciens

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002

C:Accession: F98228

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,

A.; Liu, F.; Wolism, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Matkeltz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: F98228

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-735 <KUR>

A:Cross-References: GB:AE007870; PIDN:AAK9352.1; PID:g15159198; GSPDB:GN00170

C:Genetics:

A:Gene: AGR_L1558

A:Map position: linear chromosome

C:Superfamily: 1,4-alpha-glucan branching enzyme

Query Match 3.3%; Score 124.5; DB 2; Length 735;

Best Local Similarity 20.4%; Pred. No. 3.9;

Matches 157; Conservative 83; Mismatches 255; Indels 275; Gaps 44;

```

Qy 108 DGFVSKTM---GELPPIPLSVKYLKIKYVTDKVGOLKGTDEMERLRKAQVCOELA 164
Db 43 EGSARCFIPGAEVSVLTLDGNFVGLKQIDP--DGFEGRIDLSKRPPVRYRACR--- 97
Qy 165 KRENADENOLCIVEPEYSN-----NEVF-----LIYDVFEDVYM----- 199
Db 98 -----DDAEWAVYDPSFSPVLGPMDDYFVRBSGSHRLTFD-----RMGAHPLKLBEGVEGF 147
Qy 200 ---VFAPPSSVGKFGGDTDNMMMPRH-----TGDFSVFR--VYAGADNRPAYEYSKDNK 247
Db 148 HFAVMAFNARAVSVYGDFFNNMDSRRHVMFRKDTGIMEIFADPVYAGC----- 195
Qy 248 PIKPYVFAAVSMOG---YKADYA-----MTIGPGSTDRYLLTSGWEDRIENEN 294
Db 196 AVK---FEILGANGELLPKADPYARGBELRPKNASVTAPELTKQW-----EDQAHREH 246
Qy 295 NRRIEVR-----GIKQIWKEA---MSADQ--ATIKYASK----- 325
Db 247 WAQVDRORPISITYEYHNASQWRREDGTFLSWDELAQOLIPYCTDGMGFTHIEFLPTEHP 306

```

```

Qy 326 -----YAOSANYMKNISIGNRGLARLDYIGRKAERARADMT----- 363
Db 307 YDPSMGYQFTGLYAPYARF-----GDPEGFARF--VNGAKKVGIGVLDDVPAHPPTDEHG 360
Qy 364 -----RKNGKSAVYG---DVYSSL-----EKAYKEGAK--AN 390
Db 361 LRMFDGTALYEHADPRQGFHPDMNTAIYNFGRLEVMSYLIINNALYWAKEFHLDGLRVDAV 420
Qy 391 REMYLSLETL-----FGG---TEVYRFAQFAALATNDPAHAGILKSLDDKDYDL 438
Db 421 ASMLYLDYSRKGEWIPNTEYGGRENLESYFLQKNSLYYG--THFGVTTAEBSY--W 476
Qy 439 PSIDRRKVLPRAMLDIYRRIRPADKLPDIFKNVIDKKRKGDTKRYADVFDPKSVVPYSDKH 498
Db 477 PKYSQPVHEGGIG-----FGFKNMGMFMHDTLSY---FSRE--PVHRRKF 516
Qy 499 AM-----LKSMDKEKRAKAIKEDPAVELSKSYIAAARAIQADAMANAAYAI-----EKKK 547
Db 517 HOELTFGLLYAFETNFVPLSHDEVVHGKSLAKNSGDDMOCKFANLRSYSGFMMGYPCK 576
Qy 548 RLFFAG-----LRE--MYPG--RALPSPDANTMMSYG-SIKGYEPQD 585
Db 577 KLLFNGQEPFAQMSSEMSKESLDNMLRQIYPMHGEGRRLVLDNLITYSKAALHARDCPEPG 636
Qy 586 GAM--YNYHTTGKGVLEKODPKSDEFRAVOENILDFRTKNY-----GRYAENGQLHIAF 637
Db 637 FRMLVYDDHENSVFAMLRTPAPEGKPAVICNLTPYRENVYVPLPVAGHMR----- 688
Qy 638 LSNND--ITGNSGSPVPFDKNGRLIGLAFDGNMEAN-----SGDIEFEED 680
Db 689 ILNTDAEITYG--SGK---GNGGRVQAVDAGGELGALVLPPLATIMLEPE 734

```

RESULT 7

AH3057

glycogen branching enzyme glbB [imported] - Agrobacterium tumefaciens (strain C58, Du

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002

C:Accession: AH3057

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Moks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; McCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perty, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AH3057

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-735 <KUR>

A:Cross-References: GB:AE008689; PIDN:AA144878.1; PID:g17742527; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: glbB

A:Map position: linear chromosome

C:Superfamily: 1,4-alpha-glucan branching enzyme

Query Match 3.3%; Score 124.5; DB 2; Length 735;

Best Local Similarity 20.4%; Pred. No. 3.9;

Matches 157; Conservative 83; Mismatches 255; Indels 275; Gaps 44;

```

Qy 108 DGFVSKTM---GELPPIPLSVKYLKIKYVTDKVGOLKGTDEMERLRKAQVCOELA 164
Db 43 EGSARCFIPGAEVSVLTLDGNFVGLKQIDP--DGFEGRIDLSKRPPVRYRACR--- 97
Qy 165 KRENADENOLCIVEPEYSN-----NEVF-----LIYDVFEDVYM----- 199
Db 98 -----DDAEWAVYDPSFSPVLGPMDDYFVRBSGSHRLTFD-----RMGAHPLKLBEGVEGF 147
Qy 200 ---VFAPPSSVGKFGGDTDNMMMPRH-----TGDFSVFR--VYAGADNRPAYEYSKDNK 247
Db 148 HFAVMAFNARAVSVYGDFFNNMDSRRHVMFRKDTGIMEIFADPVYAGC----- 195

```

```

OY 248 PYKVYVYAAVSMOG-----YAADYA-----MTIGPGSTDXYKLSWGEDPIEMEN 294
Db 196 AYK---FEILGANGELLPLADYARRGELRPKNASVTPARPELTOKW-----EDQANREH 246
OY 295 NPRIEVR-----GIKOGYIMKKA-----MSADO--ATRIKASK-----325
Db 247 MAQVDQRQRPSTIYEVASWQMRREGDTLSWDELAAOLIPCTDMGFTHTIEFLPTEHR 306
OY 326 -----YAOSANYWKNKSGMNRGLARDVYIGRRRAERAFADVI-----363
Db 307 YDPSMGYOTTGLYAPARF-----GDEPGRAR--VNGAKHYGIGVLLDMVPAHPTDEHG 360
OY 364 -----KKNKSAVYG-----DVLSL-----EKAYEGAK--AN 390
Db 361 LRWEDGTALYEHADPRQGFHPDMWNTAIYNGRIEVSUYNLNAALYAEKFIHLDGRLADV 420
OY 391 REMTYLSETL-----FGG---TEVYRQAFNALATNPDAHGILKSLDDXYKDYL 438
Db 421 ASMLYLDYSRKEGWIPINEJGGRNEJESVRFLOKANSIYLG---THPGVTTIAEESTS--W 476
OY 439 PSIDRKVLPAMLDIVRRRRIPADKLPDIFKNVLDKKKKGTKKYADFEVFDKSVVPYSDKH 498
Db 477 PKVQPVHEGGLG-----FGKMMGMGMDTLTSY-----PSRE--PVHRRFH 516
OY 499 AM-----LKSMDKEFKPAKIEKDPRAVELKSXYLAARAIQADAMANAIAI-----EKGK 547
Db 517 HOELTFGLIYAFTENFVLPISHDEVVHKGKSLAKSGSDMOKFANLRSKYFGFMWYGGK 576
OY 548 RLFPAG-----LRE--MYP--RALPSPDAFTYRMYSYG--SIKGYEPD 585
Db 577 KLEMGGEFAOMSEMSKSGSLDMWLNQYPRHESMRRLVLDLITRKSAAHLARCPEDG 636
OY 586 GAW--YNYHTTGKGVLEKODPKSDEFAVOENILDERFTNY-----GRYAENGOLHIAF 637
Db 637 FRMLVVDHENSYFAMLRTPARBEKPAVAVICNLTPVYRENYVPLDPAVGWRE-----688
OY 638 LSNND---ITGNGSGSPYEDKNGRLIGLADGNNEM-----SGDIEEPD 680
Db 689 ILNDAIATYGG--SGK---GNGGFRQVAVDAGGEIGALVLPPLATIMLEDE 734

```



```

      302 VSKYEGSKGADLVLSYDFEOKAVEDILKRNK-----GQKQVAGSD 343
      485 VEDKSVY-----PYSDFHML--KSMDEKFAKAIKDPAVELSKVIAARAIOAD--- 535
      344 LEPRAEVVAMPDPSGCVLALAGOKLNDKEF-----DDSYLGFETAYAMGSAVKSTLL 398
      536 -----AMAN-----ATAIEKGRLEF-----AGLEMYGRALPBDAN--FTMR 573
      399 GGIMGCAISNKVYFTDQPIALKGTKPKSSWFENRTGANKPLDVGALETISSNYSMTQVAM 458
      574 SVGSIKGYEPO-----DGAMYNHTTGKGYL-----EKODPKSDEFAVOENIL 616
      459 KMGGA-KYVNGPLRAPLSTPDMRYYNQFGLGVKTGIDLPBGQYKGGDOOTICK-IL 516
      617 DLFRTNRYGAYENGQLIHAFLSNNDITGNSGSPVFDK-----NGRLIGLAFDGNME 669
      517 DF-----AIGQYSYTPLOMAQVYSTIANGSGSRAPSMVEKIRNPSTNGDSTLATAN-- 570
      670 AMSGDIEFPDLORTISY---DIRYV 692
      571 -----EPKVLNKIGVSDNDIKTY 588

```

RESULT 12

```

D82883
DNA helicase II U0501 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: D82883
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Helner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to Genbank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum. Alternate views of a mit
A:Reference number: A82870
A:Accession: D82883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-743 <GLA>
A:Cross-references: GB:AE002148; GB:AF222894; NID:g6899495; PIDN:AAF0913.1; GSPDB:GN001
A:Experimental source: serovar 3, biovar 1
C:Genetics:
A:Gene: uvrD; U0501
A:Genetic code: SGC3
C:Superfamily: helicase II

```

```

Query Match          3.3%; Score 122; DB 2; Length 743;
Best Local Similarity 17.7%; Pred. No. 5.7;
Matches 143; Conservative 129; Mismatches 281; Indels 256; Gaps 40;

```

```

      10 LGAAALLGASGVAKAD-----KGMMLNELNOENLDMRELGFLL--PLDSLXSPDKPSI 62
      22 LGQQLVIAAGNGTGTIVLRLAYLITE--KNHPSRLIGFTTKKADDEM---KERV 74
      63 ANAVNV-----FGGCTGTVSDOGLIFTNHHCYGAIQOSYVDHDIYLRDGFVSRTM 115
      75 GKTIGVSIPYLSFHSMC--VKILQDHYLNVH---NNIKIIDDOEVL-----LKEI 124
      116 GEEPLPIPGISVATKRIKIVYTKVBSQGLKGLTD-----EMERLKAQGV----- 159
      125 FQDLNTEKRS-QVTKIKIITISKVKNKFPQNDMLNEKNHKYLELDVLDNAQRLVDIYKI 183
      160 -COELAKKENADENOLC-----IVEFYSNNEFLLVYVFXDVR----- 198
      184 YDCRCERLNVDDDLINLTHKLFIETPEYLERMOKNKEDYILV--DEFQDTNKIQYDLIS 241
      199 -----MFAPPSSVGKFGGDTDNMMMPRHGTGDSFYFVYAGADNRPAYSKDN-K 247
      242 LATKHNQNLVVGDDPDQMIYSFNG-AEQWIIINNFQSNFK-----NTKYVILITNTR 291
      248 PYKPYTFAASMOGYKADYAMTIGFPGSTDRYLTSMGVEDRIENENNPRIEVRG---IK 304
      292 STQPIINTANRLIDANNNNY-----KKMLTAFNT-----NDNNLPITYLRGQNPID 336

```

```

      305 OGIM-----KEAMSADQATRIKASKYAOSANYMKNISGMNGLARLDVIGKRAE 355
      337 EAMWIAKRTRELEESTPANOIAVLFPSNHYSITIQ-----SMARESPITILSGKRY 391
      356 ERA-FADWIRKNGKSAVYDVLSTLEKAYEGAKANREMYL-----SETLFGTEVVR 409
      392 ERAEIKDMI-----AYLKVNDLD-----ELSPRIINTPRRAIGPTFEHV 433
      410 AQFA-----NALATNDAP-----AGILKSDDKKDYLPISLDRVLEPA 448
      434 KHAIINNNLFEALAEVERKNHILNNTQKKNILNFVNLIKRIEDEMDL-----KIME 486
      449 MLDIVARRI-----PADKLPIEFKNVIDKFEKGTGRKVAQFVFPKSVVPSDKPHML 501
      487 ILELYKKVYENAYLLENKAEDKIDNYE--LKRMMKYVDRHPDITINDLINSTALVL 544
      502 KSMDEK-----FAKAIEKDP-AVELSKVIAARAIOADAMANAAYIEKRL 549
      545 NKDGKSKENVLLMTVHNSKGLEYNVFAVAGMEGLLPDRAINDDPK---GVEEERRI 601
      550 FFRAGL-----REMYPRALPBDANFTYRMASGSIKGYEPQDAMY 589
      602 AYVALTRAKKNLIYSSACCYDPLARQVESPRIENIGFNKNLIINSFKNKPED--- 657
      590 NYHTTGKYLEKODPKS--DEFAVOENILDFRTKNYGRYAENGQLIHAFLSNNDITGCG 647
      658 ---MPLKSLFKQDEERSWPDSDKOKOKEVEDNFIQ-----TKND----- 693
      648 GSGSPVFDKNGRLIGLAF-DGNMEAMSGDI 675
      694 -----FEIGERIVHTSFGDGVIIIGDGI 717

```

RESULT 13

```

C95066
valyl-tRNA synthetase [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95066
R:Rietveld, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Mayhew, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; M01D:2157209; PMID:11463916.
A:Accession: C95066
A:Molecule type: DNA
A:Status: preliminary
A:Cross-references: GB:AE005672; PIDN:AAK74724.1; PID:q14972043; GSPDB:GN0164; TIGR:
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0568
C:Superfamily: valine-tRNA ligase

```

```

Query Match          3.3%; Score 122; DB 2; Length 883;
Best Local Similarity 20.0%; Pred. No. 7.4;
Matches 102; Conservative 66; Mismatches 167; Indels 176; Gaps 24;

```

```

      192 DVF-----KDVNVFAPPSVGKFGGDTDNMMMPRHGTGDSFYFVYAGADNRPAYSK 243
      26 DVFKPSGDOAKAKYSTIYPPNVTGKL-----HLG-----HA 57
      244 KDNKPKPYFAVAVSMOGRKADYAMTIGFPGSTDRYL-TSMGVEDRIENENNPRIEVRG 302
      58 WDT-TLQDIIIRQKRMQGD-----TLWLPQMDHAGIAFQAKVEERLGEIGTFRDLCR 110
      303 -----IKQGIKEMASADQATRIKYA-SKYAQSANTYKNKNSIGMNGRLARLDVIGRAAEER 357
      111 ESFLTFVWEWKD-----EYATTIKEQWGMKGLSDVYSRERFTIDEGIS-----KAVRK 158
      358 AFADWIRK-----NGKSAVYGD-----VLSSLEKAYK-----EGAKANREMT 394

```

```

Db      159 VFVNLYKKGMWYIRGSEFIINMPDAARTALSDIEVITHKVDGFAFYHNNMYMLEBGSRRLEAVT 218
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      355 YLSEFLFEGTEVYVRAQAFAMALATNPDAAHAGITLSDKKYDYLPs-----LDRKVLPRAM 449
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      219 TRPETMFGDV-----AVAVNPE-----DERYKDLIGKNYILPIANKLPIV 259
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      450 LD-----IVRRRIPADKLPIDFKNYIDKTKKGGPTKKADVVPKSVYPSYSDK 496
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      260 GDEHADPELGCGVUKITPAHDPNDFLVGGRHNLPVNYMMDGTGNNELAFESGM--DR 316
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      497 FHAMLKSMDKERFAKAIKEKDPAVELSKVSIAARA-----IOAD 535
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      317 FEAR-----KAVVAALAEELGALYKIEKRHSYGHSERTGVVVERPLSTQWFKMDQLAKN 371
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      536 AMAANVAIEKKRFLFFAGLRREMYRGRALPS-----DANFTYKMSYGSITKIEPDG 586
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      372 AIANDOTEDK-----VEEYPRPFNDTFLQWMEVNHWDWISKQLWMGH-----QIP 416
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      587 AMYNHTTGKGVLEKODPKSDEFQAVQENILD 617
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      417 AMYN--ADGEMTVGEAPEGDGTQDEDEVLD 445
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

RESULT 14
D69103
DNA helicase (EC 3.6.1.-) MTH1770 [validated] - Methanobacterium thermoautotrophicum (strain ATCC 35061) [NCBI]
N:Alternate names: cell division control protein CDC21/CDC54 homolog; replication licens
N:Contains: adenosinetriphosphatase (EC 3.6.1.3)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 26-May-2000
C:Accession: D69103
R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicalre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jivanli, N.
J.; S. Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514
A:Accession: D69103
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-666 (MTH>
A:Cross-references: GB:AE000932; GB:AE000666; NID:2622894; PIDN:AMB86236.1; PID:q262296
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1770
C:Complex: dodecamer; double hexamer [validated, MUID:20144074]
A:Function: <HEL>
A:Description: EC 3.6.1.-; DNA helicase; ATP-dependent DNA unwinding activity [validated
C:Function: <ATP>
A:Description: EC 3.6.1.3 [validated, MUID:20144074]; adenosinetriphosphatase, DNA-depend
C:Function: <DNA>
A:Description: ATP-independent single-stranded DNA-binding, ATP-dependent double-stranded
C:Superfamily: Archaeoglobus replication licensing factor MCM-type homolog, MCM homology
C:Keywords: cell cycle control; DNA replication initiation; hydrolase
i:98-570/Domain: MCM homology <MCM>

```

Query Match      3.2%; Score 120; DB 1; Length 666;
Best Local Similarity 19.3%; Pred. No. 6.6;
Matches 116; Conservative 94; Mismatches 232; Indels 158; Gaps 28;

QY 119 LPPLGLSVKYLEK-----IVKPTDKEVGQL-----KGIIDEMERLKAQEV 159
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 96 IPLRLRLRSKFLICKFAVVDGIYAKTDEIRPRIVKANFEGRCGCMRHIAVQSTNMITE-PSL 154
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 160 COELA-----KKNADENQLCIVPEF--YSNNEXFLIIVDYDKDVMFAEPSSVG 208
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 155 CSECGSRFSRLDQDESEFLDTQTLKQPLENLSSGEGPRLTVLVLEDDLVDTLPGDIIV 214
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 209 KFGGDT-----DNMMMPKHT-----GDFSIFRYRYAGADNRPARYSKDNKPY-KPV 252
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 215 RYTGILRTVARDERTKRFKNFIVGNTYLFELQEFEEFLQISEDEEKIKLADDPNIYEKII 274
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

OY      253 YFAASMOGYADVAMITIGFPGSTDFYLTSWGVEDRIENENNPNRIEVRGKIOGIMEAM 312
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      275 RSTASHINGREYVKALILQLEFGTGKEL-----DCKRLRGDIHILIIV--DPGIGSQM 328

OY      313 SADOATRIRKYASKYAOSANYWK---NSIGNRGLARLDVIGRRRAERAFADMIIRNGK 368
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      329 -----LKYYSKLAPRGIIYTSKGCTSGVGLTAAAVRDE-----FGGSLEAG- 369

OY      369 SAYVD-----VLSSLEKRYKEGAKANREMTYLTSETIEGTETVYRFAQFANMLATNPDAHA 424
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      370 ALVLIDKGNVCVDLDMKKREEDRSAIHE-----ALEQQTISIKA 409

OY      425 GILKSILD-----DKYKDYLPSLDKVLPA-----MLDIVRRIRIPADK 461
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      410 GIMATLNRSVCYLLAANKRCGRFDYSKIAIOID--LPRTILSRFLDIIPEVEKRPDEEK 466

OY      462 LPDIFKNVIDKKFKGDTKRYADVFWDKSVPEYSDKFHAMLSMDKEKPAKAIIEKDPAVEL 521
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      467 DRELRLNR-KTHKED---HMPEFELDPELL--RKVIYAR-----KNVRPVLIDEAMQV 514

OY      522 SKSYVAARALIOAD----AMANAYAIIEGKRULFRAGLBREMPCGRALPSBANFTMRMSGS 577
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      515 LEDFYVSBRASAADSDSPVITAROLEALVLRSESSAIKLKEHVEADEAKARKILSQC 574

OY      578 IK--GEYBODGAWMYHTTGKVLEK---ODPKS--DEFAVOENILDLFFRNKNGRAEN 630
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      575 LKQVYDE-----TGKIIDDKVEGRPKSRBDKFULLLEIKETYEDDYGGRAPTN 625

```

RESULT 15
CB2332
translation elongation factor EF-G VC0361 [Imported] - Vibrio cholerae (strain N16961)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_Revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: CB2332
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
Charidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
A:Accession: CB2332
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-698 <HEI>
A:Cross-references: GB:AE004124; GB:AE003852; NID:96654770; PIDN:AAF93534.1; GSPDB:GM
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0361
A:Map position: 1
C:Superfamily: translation elongation factor G; translation elongation factor Tu homo
E:11-145/Domain: translation elongation factor Tu homology <EU>

	Query Match	3.2%	Score 118;	DB 2:	Length 698;	
	Best Local Similarity	18.4%;	Pred. No.	9.5;		
	Matches 120;	Conservative 83;	Mismatches 202;	Indels 246;	Gaps 26;	
QY	8 ILLGALLL-GASVAKADKGM-----LLNELNQENLDRLRELGLTLPDLSL	54				
	: :: :: :	:	:::: :			
Db	105 VLDGAVVVFCGTSGVEPSETVMROADKYGVBMFVNKKMRACAGDLRVVG----	156				
QY	55 YEFDPRTANNAVY----FGCGCTGITV-----SDQLIEFT-----	86				
	:: :: :	:	:::: :			
Db	157 -QIKHRLGANPPIDNLNGAEBEFEKGVIDLIKMAINNEVDQGSEFYEEIPADMLELA	215				
QY	87 ---NHHCYGAIQSQTVDHYDLRDGFVS-----RTMGELPIPGLSVAYLRKIIV	133				
	:::: :: :	:	:::: :			
Db	216 QEMRHHLVFAAAEABEELMEKYLEDEGDSELSEVIKQALRQRIINNEYIYLACGAFAFKNGV	275				
QY	134 K-VTRYK-----EQOLKITDMERLKRADEVCOELAKKEADENOQICTEPFPSN	183				
	: :: :: :	:	:::: :			
Db	276 QAVLDVAIVEFLSPFDVPAIKIGIDD-----RENSVERRADQN---EPF--S	316				

```

QY 184 NEYFLIYDVFKDVRMVEAPSSVKGFGDPTDMMWPRHTGDSVFVRVYAGADNRPPEYS 243
Db 317 SLAFKTIATDF-----VGSUFTIRVYSGVYNSGDVY 348
QY 244 KDNKPYKPYEAASVSGYKAD-----DYMTIG----- 272
Db 349 NSYKOKKEREGRIVQMHANKRDEIKELRAGDIAAIGLKDVTTGDTLCDPNHVILERME 408
QY 273 FPGSTRYLTSWGVEDRIENENNRIEVRIGIKGIMKEMSADQATRIYASKYAS-- 329
Db 409 EP-----EPVIOIAVEPRSKADOEKMGIALGKLAEDPSFRVETDAETGOTLI 456
QY 330 -----ANYMKNSIGMNRGLARLDVIGRK--RAEERAFADNIRKNGKSAVYG--- 373
Db 457 SCMGELHLDIYDRMKRREGEVDCNVGKPOVAYARETIRGKSEVEGKRVQSGRGQYGHVW 516
QY 374 -----DVLSSLEKAYKEGAKANREMTYLSSETLFGGT 404
Db 517 LKIEPAEPGQGFVVDIAAGVLPKEFINPVAKGIEQNMNNGVLAGYPVLDVKATLFDGS 576
QY 405 -EYVREAOFANALATNPDAHAGILKSL-----DDKYKDYLPsIDRK--VL 446
Db 577 FHDVDSSEMAFKTAGSMAFRKGALEAPVLLPELPMKVEITTPEDMMDGVVDLNRRTGII 636
QY 447 -----PAMLDIVRRRIPADKLDPDFKNVIDRK-----FKGDTKKYAD 483
Db 637 EGMDEGPAGIKTIHAKVP---LSEMGYATDLRSATQGRASYSMERAFETAD 684

```

Search completed: October 17, 2002, 22:40:19
 Job time : 49 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 21:37:45 ; Search time 25 Seconds

(without alignments)
1102.733 Million cell updates/sec

Title: US-10-008-355-2

Sequence: 3719

1 MQMKLKSILGALLGASG.....LFMDKMGCCPRLLIQLKLI 712

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	123.5	3.3	776	YL4_CAEEL	Q05036 caenorhabd
2	122	3.3	1104	COLA_CLOPE	P43153 clostridium
3	121	3.3	1628	NAGH_CLOPE	P26831 clostridium
4	116	3.1	952	TOPI_STRCO	Q9X909 streptomyc
5	114	3.1	620	DNK_PORPU	P30723 porphyra pu
6	114	3.1	734	GLGB_AGR7	P52979 agrobacteri
7	114	3.1	747	VIVC_BPT7	P03725 bacterioph
8	113	3.0	1047	RIRI_CHIMU	Q9P193 chlamydia m
9	112.5	3.0	657	CN16_HAEIN	P44764 haemophilus
10	112	3.0	556	EST2_CAEEL	Q07085 caenorhabd
11	112	3.0	1504	DPO2_YEAST	P14284 saccharomyc
12	111.5	3.0	878	ACON_RICPR	Q9ZC14 rickettsia
13	111.5	3.0	922	YK16_CAEEL	P42173 caenorhabd
14	111.5	3.0	2410	MOXI_SCHPO	Q9U8K8 schizosacch
15	110.5	3.0	1039	SYI_MERJA	Q58357 methanococ
16	109.5	2.9	768	PARC_NEIGO	P48374 neisseria g
17	109	2.9	681	MP10_HUMAN	Q00566 homo sapien
18	109	2.9	3672	LMZ2_CAEEL	Q21313 caenorhabd
19	108.5	2.9	507	YV42_CAEEL	Q18416 caenorhabd
20	108	2.9	638	NEC2_PIG	Q03333 sus scrofa
21	108	2.9	1341	RPAL_METJA	Q58445 methanococ
22	106.5	2.9	747	Y030_UREPA	Q9P195 ureaplasma
23	106	2.9	1592	GTF2_STRDO	P27470 streptococ
24	105	2.8	1023	HLV1_ECOLI	P09983 escherichia
25	105	2.8	1391	RPOB_MYCPN	P78013 mycoplasma
26	104	2.8	616	ACOC_SOLTU	Q04916 solanum tub
27	104	2.8	908	H104_YEAST	P31539 saccharomyc
28	104	2.8	1060	DP3A_LACLA	Q9J170 laccococcus
29	104	2.8	4563	APB_HUMAN	P44114 homo sapien
30	103	2.8	320	PTA_MYCGE	P47541 mycoplasma
31	103	2.8	843	PUL4_THEMA	Q33840 thermocoga
32	103	2.8	1131	APCE_ANASP	P80559 anabaena sp
33	102.5	2.8	396	GTS1_YEAST	P40956 saccharomyc

34	102.5	2.8	718	1	FLGE_HELPJ	Q9ZKY0 helicobacte
35	102.5	2.8	798	1	UNR_PAT	P18295 rattus norv
36	102	2.7	555	1	MCRA_METPE	P12971 methanother
37	102	2.7	560	1	YDRN_ECOLI	P77318 escherichia
38	102	2.7	626	1	HTPG_BACSU	P46208 bacillus su
39	102	2.7	656	1	HTPG_MYCLE	Q03012 mycobacteri
40	102	2.7	742	1	KM65_YEAST	Q03566 saccharomyc
41	101.5	2.7	627	1	CACP_CANTR	Q00614 candida tro
42	101.5	2.7	629	1	KSKY_MOUSE	P48025 mus musculu
43	101.5	2.7	675	1	RG59_MOUSE	Q54828 mus musculu
44	101.5	2.7	770	1	DBP4_YEAST	P20448 saccharomyc
45	101	2.7	500	1	Y039_BORBU	O51068 borrelia bu

ALIGNMENTS

RESULT 1	ID	YL4_CAEEL	STANDARD:	PRT:	776 AA.
AC	005036	YL4_CAEEL	STANDARD:	PRT:	776 AA.
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Hypothetical 86.9 kDa protein C30C11.4 in chromosome III.				
GN	C30C11.4				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;				
OC	Rhabditidae; Peloderinae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BRISTOL N2.				
RX	MEDLINE=94150718; PubMed=7906398;				
RA	Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,				
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,				
RA	Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,				
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,				
RA	Johnston L., Jones M., Kershaw J., Kristen J., Laister N.,				
RA	Latteille P., Lightning J., Lloyd C., Mottimore B., O'Callaghan M.,				
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,				
RA	Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,				
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,				
RA	Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,				
RA	Wohlman P.,				
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.				
RL	Nature 368:32-38(1994).				
CC	-1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY. STRONG,				
CC	TO YEAST MS13				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement. (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch)				
DR	EMBL: L09634; AAA27967.1; -				
DR	PIR: S44784; S44784.				
DR	HSSP: P19120; 1NCG.				
DR	WormBep: C30C11.4; CEB00103.				
DR	InterPro: IPR001023; HSP70.				
DR	Pfam: PF00012; HSP70; 1.				
DR	PROSITE: PS00297; HSP70_1; FALSE_NEG.				
DR	PROSITE: PS00329; HSP70_2; 1.				
DR	PROSITE: PS01036; HSP70_3; 1.				
KW	Hypothetical protein: ATP-binding; Multigene family.				
SQ	SEQUENCE 776 AA; 86896 MW; 20FA975FB649FA9C CR64;				
Query Match	3.3%;	Score 123.5;	DB 1;	Length 776;	
Best Local Similarity	21.1%;	Pred. No. 2.2;			


```
OY 589 YVHTTGKGVLEKODPKSDEFVQENIIDLFRTKNYG-----RYAENGQLIHT----- 635
DB 607 ANY-SEAGAFERK-----SKTGYGFHYVDHTEXAEVGVQHIYPIKLS 646
OY 636 -----AFLSN-NDITGNSGSPVEDKN 656
DB 647 MCONLSVIGSTIVDPNRITATYISNRDAPGPNPDN-IFDNN 687

RESULT 4
TOP1_STRCO STANDARD; PRT; 952 AA.
AC 09X909;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (unwisting enzyme) (Swivelase).
GN TOPA OR SCH5.06C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D., Bentley S.D., Parkhill J., Barrell B.G.,
RA Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -I- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -I- SUBUNIT: MONOMER (BY SIMILARITY).
CC -I- MICELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -I- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL035636; CAB38480.1; -.
DR HSRP; P06612; 1ECL.
DR InterPro; IPR003601; DNATOP1_ATP_bind.
DR InterPro; IPR003602; DNATOP1_DNA_bind.
DR InterPro; IPR000380; Pro_topoisomerase.
DR InterPro; IPR002936; Toprim.
DR Pfam; PF01131; Topoisom_bac; 1.
DR Pfam; PF01751; Toprim; 1.
DR PRINTS; PR00417; PRTPI5MRASEI.
DR SMART; SM00437; TOP1AC; 1.
DR SMART; SM00436; TOP1BC; 1.
DR SMART; SM00483; TOP1IM; 1.
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW Isomerase; Topoisomerase; DNA-binding.
FT ACCT SITE 334 334 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 952 AA; 103584 MW; 1C98DABD7A6292925 CRC64;
```

```
OY 156 AQEVCQE-----LAKKENADERNOLC-----IVEPYSNNEFLLIYVDKX-VRKVFAP 204
DB 89 LKDLKESDELFLATDDEDEGBEALAMHQLQEVLPKRIPIYKR--VWFHEITKAIIRAIVANP 146
OY 205 SSVGK-----FGDIDNNMMPRHIGDSVFVFYAGADNRPAPEYSKDKPKY 249
DB 147 RELNOKLVDQAQETRRILDRILRYLGYEVSPLYMKVMPRLSAGVQSVATPLVVEREREIAF 206
OY 250 KPYVEAAVSMOGYKADVDYAMTIGPFGSDRYLTSVGVEDRIENENNPRIEVRGIKOGIWK 309
DB 207 RSAEY--WDLTGTFRATGRAGASDPSSLVARIQT-----VDRRVAQG--- 247
OY 310 EAMSADQATRIKAYSK-YAOSANYWKNISIGNRGIARLDVIGKRAEERAPADW----- 362
DB 248 --RQDSDISGLKSNANTLHLDENARALAAALENTRFANRVSSEKPYRRSPAPARTTTLQ 305
OY 363 ---IRKNGKSAVYGDVUSLSLEKATKEG--AKANREMYLSETLFGGTEV-VRFAPANAL 416
DB 306 QEASRKLGFGA--KSTMQVQAQKLYENGITYTMRDSTTLSDTAVSAAARQVTOLYGADYL 363
OY 417 ATNPDAHAGILKSLDDKXKDYLPISLDRKVLPA-----MLDIYRRRIAPADKLPDI 466
DB 364 PPQPTTYAGKVKNAQEAHEAIRPSGDRRTPEAETGLGDOFKLYELMKRTVASOMKDAT 423
OY 467 KNVIDKRFKGGDTKKYADVFDFKSVVYPYSDKFFHMLKSM-----DKERFAKAI- 513
DB 424 GNSVYKVIKGAASDGRDVEFASGKTI--FHGFLKAAVEGADDPNMLDREHRLPQVA 481
OY 514 -----EKDPAVELSKSYI-----AAARAIOADAMANAVALIEKG 546
DB 482 EGDALTAETIIVDGHATPAPARTYEAISLYKELEREIGRPSTYASITIGTIIIDRGVFKG 541
OY 547 KRLF-----FA--GLREYPPARALPSDANFTMRM--SYGSIKGYEPODGANYNH----- 592
DB 542 TALVPSTLSFAVNVNLLEHFRFLV--DDEFTARMEDDLDIRIARGAQSVPLRRPYGEG 599
OY 593 --TTGKGVLEKODPKSDEF-AVENIIDL-----FRTANYRGYENG 631
DB 600 DGTGGGGAADNGDGHDLGLGKELVTDLGAIDAREVSSFPVGNDIKRVGRPYPYERG 659

OY 632 Q 632
DB 660 E 660

RESULT 5
ID DNAK_PORPU STANDARD; PRT; 620 AA.
AC P30723;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chaperone protein dnaK (heat shock protein 70) (heat shock 70 kDa
DE protein) (HSP70).
GN DNAK.
OS Porphyra purpurea, and
OS Porphyra umbilicalis (Laver).
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxID=2787, 2786;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=P.purpurea; STRAIN=AVONPORT;
RA Reith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=P.umbilicalis; STRAIN=AVONPORT;
RA Reith M., Munholland J.;
```



```

RT "An hsp70 homolog is encoded on the plastid genome of the red alga,
RT Forphyra umbilicalis."
RL FEBS Lett. 294:116-120(1991).
CC -1- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U38804; AAC08201.1; -
DR EMBL: X62240; CAA44160.1; -
DR PIR: S19660; S19660.
DR HSSP: P04475; 1D64.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
DR Chapterone; ATP-binding; Heat shock; Chloroplast.
SQ SEQUENCE 620 AA; 67619 MW; C9FB4713C142FECEP CRC64;

Query Match 3.1%; Score 114; DB 1; Length 620;
Best Local Similarity 19.8%; Pred. No. 6.4; Indels 198; Gaps 32;
Matches 128; Conservative 82; Mismatches 198; Gaps 32;

QY 51 LDSLVSFDRKPSINAVVIFGGCTGIVSDGLIFTNHHCY-----GAIQSSTVD 102
DB 1 MGRVVGIDLGFTNSVIAVMGGRKPTVAPNAGGRITASVAAYIRKSGDKLVQIARQAVIN 60
QY 103 HD---YLRDGFVRITGE---ELPIPELSYKRYRKIVYTDKYGOLKITTDEMERIRK- 155
DB 61 PENTTFYSVKRFFIKRQNEISQEIROTSYNVTKSGSSIKI-----ECPALMKD 107
QY 156 --AOEQCOELAKK--ENADE-----NOLCIPEFYSNNRYFLIYVDVK---DY-RMV 200
DB 108 FAEELISAQVLRKLVDASTYLGELYTQAVITTPATFNDORCATDAGKIAGIDVLRIT 167
QY 201 FAPPSVGRKGGDTN-----MMMPRHGTDFSVFRVYAGADNRPAEYSKDNKPYKVPV 254
DB 168 NEPTAASLSTYGLDKONNETILVFDLGGGTFDVSILEYGDV-----F 209
QY 255 AAVSMOG---YKADDAVMTIGFPGSTDRYLTSGVEDRIENNNPRIEVRGICQGIWKEA 311
DB 210 EVLSTSGDTHLGGDDP-----DQCIIVEMLIKDFKQSE-----GIDLGRKQRA 251
QY 312 MSADQATRIYASKYA--OSANYWKNSTIGMNRGLARLDVIGRRRAERAFADMTIRKNKGS 369
DB 252 LQ-----RLTEASEKAKIELSNLTQTEINLPFTATOD--GPRHLE-----KTVTR 295
QY 370 AAVGDVLSL-----EKAYKEGAKANREMYLSETLP--GGEVVRFAQFANAL--- 416
DB 296 AKPEELCSRLIDKCSIPVNNALMD--AKLEASSIDEVVLVGSGSTRIRPAIQQWVKRLIG 352
QY 417 ----ATNPD-----AAAGILKSLDDKYKDYLPISIDRVLPAMDIVRRRIIPADKLDP 464
DB 353 DPQNSVDPDEVAIGAAGVAGLAG---EVKDLI-----LIDVTPLSLGVEVLGG 399
QY 465 IFNNVNDKKRGKDTKKKADVFPDKSVVPYSDKFAHMLKSMDEKFAAIKEDPAVELSKS 524
DB 400 VMKTIIPRNTTIPTKK-----SEVFSTAVNQPWEL--Q 432
QY 525 VIAAARAIQADAMANAIAIEKGRLEFAGLRREMPG--RALPS-----DANFTMRSYG 576
DB 433 VLQGERELTMD-----NKSIGTFRLDGMIPARAGVQIEVTTDIDANGL----- 477
QY 577 SIKGYEPDQAMTYNHTTGKGVLEKQDPKSDERAVQENI-LDLFRTKN 623
DB 478 SYVAKERATGKEQSTITSGASTLPKDDVERHVRKAEENFVDQKRRRN 525

```

```

RESULT 6
ID GLGB_AGRU STANDARD; PRT; 734 AA.
AC P52979;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (Glycogen branching
DE enzyme).
GN GLGB.
OS Agrobacterium tumefaciens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=358;
RN (1)
RP SEQUENCE FROM N.A.
RX STRAIN=A348;
RX MEDLINE=99069330; PubMed=9851999;
RA Ugade J.E., Lepek V., Uttaro A.D., Estrella J., Iglesias A.,
RA Ugade R.A.;
RT "Gene organization and transcription analysis of the Agrobacterium
RT tumefaciens glycogen (glg) operon: two transcripts for the single
RT phosphoglucomutase gene."
RL J. Bacteriol. 180:6557-6564(1998).
CC -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
CC glycogen.
CC -1- PATHWAY: THIRD STEP IN GLYCOGEN BIOSYNTHESIS.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AF033856; AAD03472.1; -
DR InterPro: IPR000461; Alpha_amylase.
DR InterPro: IPR004193; Isoamylase_N.
DR Pfam: PF00128; alpha-amylase_N.
DR Pfam: PF02922; isoamylase_N.1.
KW Glycogen biosynthesis; transferase; Glycosyltransferase.
FT ACT_SITE 417 417 BY SIMILARITY.
FT ACT_SITE 470 470 BY SIMILARITY.
FT ACT_SITE 538 538 BY SIMILARITY.
SQ SEQUENCE 734 AA; 83623 MW; 70A3CD5A77F31E6 CRC64;

Query Match 3.1%; Score 114; DB 1; Length 734;
Best Local Similarity 20.1%; Pred. No. 8.1;
Matches 154; Conservative 84; Mismatches 261; Indels 266; Gaps 43;

QY 108 DGEVSRFM---GHELPISGVYKRYKIVYTDKVGOLKGIIDEMERLRKAQVQELA 164
DB 43 EGFSAKCFIPGAEEVSLTLDGNFVGLKQIDP--DGFEGRIDLSKQPVRRYRCK- 97
QY 165 KRENADENOLCIYEPFESN-----NEYFLIYDV-----FKDVR-----MVF 201
DB 98 ----DDEAIVADIPYSFGLVLPMDYTFRESIGSYTGMAIRPLKLEGGVGFHRAVW 152
QY 202 APPSSVGRKFGDITDNMMWPRH-----TGDVSFR--VYAGADNRPAEYSKDNKPYKPV 252
DB 153 APNGRRVSVYGDNRNMDGRHVRFRKDTGIMELIFAPDYVACA-----YK-- 197
QY 253 YFAAVSMOG---YKADDAVMTIGFPGSTDRYLTSGVEDRIENNNPRIEVRGICQGIWKEA 311
DB 198 -FELIGANGELPLKADPVYARRGELRPNKASVYAPETQKW-----EDQAHREHMAQVD 250
QY 300 VR-----GIKQGIWKEA-----MSAD--QAVRIYASK----- 325

```

Db 251 QRRPISIEYHAGSMOSEDCGFLSMDELAEQLIPYCGDMGFTHEIEFLPTEHPDPSW 310
QY 326 -----YAOSANYKNSNGMNGIARLDVYGRKRAEERAFADWL----- 363
Db 311 GYOTTGALYAPTRARF-----GDEGEFARF-VNCAHKVYGIGVLLDWPAPHPEDHGLRMFD 364
QY 364 -----RKNGSAVYG-----DVLSSL-----EKAYKSGAK--ANREMY 395
Db 365 GYALYEHADPRGCFHPDMNTATYINGRIEYMSYLNNALYMAEKHLDGLRYDAVASHLY 424
QY 366 LSETL-----EGG-----TEVRFAPAFANALATNPDAHAGILKSLDKYKDYLPISDR 443
Db 425 LDYSRKEGEMIPNEYGRENLESVRFLOKMSNLVYG--THPGVMTIAESTS--WPKYSQ 480
QY 444 KVLPLAMLDVRRRIKPADLPDIFKNVIDKKFKGDTKKADPEYEDKSVYPSKFFAM--- 500
Db 481 PVHEGGLG-----FGFKMNGFMHDTLSTY---FSRE--PVHRKFFHOELT 520
QY 501 --LKMDEKEFAKALEKDPAVELSKSVIAAARAIOADMANAYAT-----EKGRLFEFA 552
Db 521 FELLTAFTENFVLPLSHDEVHVGKSLIAKMSGDDMOKFANLRSTYGFPMGYPGKLLFM 580
QY 553 G-----LRE--MYPG-RALPSDANFTMRMSYG-SIKGYEPDQGAN-- 588
Db 581 GGEFAQWSEWSEKSGSLDMNLBQYPMHEGMRRLVRLNLTYRSKALHARDCBPDGFRMLY 640
QY 589 YVHTTGKGVLEKOPKSDSEFAVOENIIDLPFTKNY-----GRAENQGLHIALSNND 642
Db 641 VDDHNSVFAWLRTAPGKRPVAVICNLTPYKRENYVPLGVAGNRRE-----ILMTD 692
QY 643 --ITGNGSGSPVFDKNGRLIGLAFDGNEMAM-----SGDIEFEPD 680
Db 693 AEITYG-SGK---GNGGRQAVDAGGEIGAMLYPLPLATIMEPE 733

RESULT 7
VIVC_BP77 STANDARD: PRT: 747 AA.
AC P03725:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DE Internal virion protein C.
GN 15.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like phages.
OX NCBI_TaxID=10760;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RA Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RT locations of T7 genetic elements.";
RL J. Mol. Biol. 166:477-535(1983).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01146; CA24433.1; -
DR PIR: A04351; HIBPC7.
DR PIR: S42331; S42331.
SQ SEQUENCE 747 AA; 84341 MW; 959C572B7B42C2B8 CRC64;

Query Match 3 1%; Score 114; DB 1; Length 747;
Best Local Similarity 18.8%; Pred. NO. 8.3; Indels 218; Gaps 27;
Matches 114; Conservative 86; Mismatches 190;

QY 91 GYGAIOSSTVDHDDYLRDQFVSRTWGEELPIPGLSVKYLRKIVKTVDKVEGOLKGTDEM 150
Db 26 GYRAATTOAEQPRSSLD-----TIG-----RAKAGADMTYAKEDARADLAE- 69
QY 151 ERLKRAQVCEOLA---KKNADENOLCIVEPFS-----NNEYFLIYDVFKDV 198
Db 70 ---RSNELIRLRPEQREALLNNGTLLYQDDPAMEALRVKTKGNMAAYLVVDDQYKIK 125
QY 199 M-VFAPSSVGFQGGDTNNMKPRITGDFSVFRYA---GADNRAEAKSKDNKPKPYFF 254
Db 126 EGVFTREEMEEY-----RHSRLQEGAKYVAEQGIDPEVDYQ----- 165
QY 255 AAVSQGKADYAMTIGFPGSTDRYTFWGVEDIEENNPRIEVRIGQ----- 305
Db 166 -----GNGDITTEINILYGAHDFLSQQAOKGAINMS---RVELNGVLODDPMLRRPD 216
QY 306 -----GIMKEAMSAQATRI-----KYASKYAOSANYKNSIGMNGIARLDV 348
Db 217 SADFEKYIDNGLVGALPSDAQATQLISOAFSDASSBAGADF-----LMR 263
QY 349 IGRKAERERADMTIRKNGKSAVYDVL-----SSLEKAK----- 384
Db 264 VGDKR-----VTLNGATTYRELIGEQWALAVTAORSOPETDAKINEQYRLKIN 314
QY 385 -----EGAKA-----NREMYLSETLFGTEVVRFAQFANALATNPDA 422
Db 315 SALNEDERTAMEMQGIKAELDKYQPDQMTPOKEWILISAE-----QYONQMANMTRKA 369
QY 423 HAGILKSLDDKYKDYLPISLRKRVLPAMLDIV---RRRIPADKLPDIFKNVIDKKFGDT 478
Db 370 QA---KALDDSMK-----SMNK-----LDVIDKQPKRINEMVSTDKMPVENTGEF 416
QY 479 KKYADFVDSKSVYPSDKFHAMLSMD-----KKEKFALEKDPAY----- 519
Db 417 K-----HSDMVNYANKRLAEISMDIPDGAOKAMKLYQAOSKDAEFTALGTMTD 469
QY 520 ---ELSKSVIAAARAIOADMANAYAIKGRLEFAGLREMPGRLASDANFTMRMSYG 576
Db 470 AGQESAAVINGKLPERRPPADALRLRINADPQLAL-----YDQAELEPLTMM 521
QY 577 SIKGYEPQ 584
Db 522 DKQGIDPQ 529

RESULT 8
RRL1_CHLMU STANDARD: PRT: 1047 AA.
ID RRL1_CHLMU
AC Q9PL93;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonucleoside-diphosphate reductase alpha chain (EC 1.17.4.1)
DE (Ribonucleoside reductase).
GN NRDA OR TC0214.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MOPn / N199;
RC MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gall S.R., Heidelberg J.F.,
RA Linher K., Weidman E.K., Peterson J., Utterback T., Berry J., Bass S.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -I- FUNCTION: CATALYZES THE BIOSYNTHESIS OF DEOXYRIBONUCLEOTIDES FROM
CC THE CORRESPONDING RIBONUCLEOTIDES, PRECURSORS THAT ARE NECESSARY

```

CC      FOR DNA SYNTHESIS (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside diphosphate + oxidized
CC      thioresoxin + H(2)O = ribonucleoside diphosphate + reduced
CC      thioresoxin.
CC      -1- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC      -1- SUBUNIT: Tetramer of two alpha and two beta chains.
CC      (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC      LARGE CHAIN FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE002288; AAF39086.1; -.
DR      TIGR; TC0214; -.
DR      InterPro; IPR000768; Ribonucleo_red.
DR      Pfam; PF00317; ribonuc_red_lg; 1.
DR      PRINTS; PRO1183; RIBORDTASEM1.
DR      PROSITE; PS00089; RIBORED_LARGE; 1.
KW      Oxidoreductase; DNA replication; Complete proteome.
FT      ACT SITE 458 458 BY SIMILARITY.
FT      ACT SITE 672 672 BY SIMILARITY.
FT      ACT SITE 687 687 BY SIMILARITY.
FT      SITE 1043 1043 INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN.
FT      SITE 1046 1046 INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN.
FT      SITE 1046 1046 INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN.
FT      SITE 1046 1046 INTERACTS WITH THIOREDOXIN/GLUTAREDOXIN.
SQ      SEQUENCE 1047 AA; 120035 MW; 9B3FED9BFAF817AA CRC64;

Query Match 3.0%; Score 113; DB 1; Length 1047;
Best Local Similarity 18.1%; Pred. No. 15;
Matches 111; Conservative 106; Mismatches 196; Indels 200; Gaps 32;

QY      112 SRMGELPIPLGLSVKRLKRIKIVKTDVKEGOLGIPDEMELTKAOCVCELAKEKNADE 171
DB      37 TRRIDDMPLP-----EDLENSIRSTHOV-----VKEVOKIT-----D 71
QY      172 NOLCTIVEFYSNNFEFLIVYDFEKFQRMVAPSPSSVGKFGGDDNMWPHRTGDFSEVER 231
DB      72 GQVTVVERIDMWESQLYT-NGLDV-----ARYVYVR- 104
QY      232 YAGADNPAEYSKD-----NKPYKPYVFAAVSMQGYKAD--YAMTIGFGST 277
DB      105 ----DDEKARREKSMQSLSVIRRCGTIVHFNPMPKISAALEKAFRATDRIEGMTPDFVREE 160
QY      278 DRYLTS---WGVEDRIENENNRIEVRGIGKQITWEKMSADQATRIYASKYQASANY-- 332
DB      161 VNAITQIVAEIERGCSQOQS-RIDIEQIDIDYEQQLM-----VGHYATAKVYIL 210
QY      333 WKNISGNRGLARLDVYIKRAERAFADWIRKNGSAVYGDVLSLEKAYEKAGKANRE 392
DB      211 YREA-----RARVRNRRVEDQIVEAPSEETFEVLS-----KDS----- 245
QY      393 MTYISETLFGTEVYVRAQFANALATNPDAHAGILKSLDKYKYDLSL-DRKVLPAMLD 451
DB      246 -TYMI-----THSQLARLARACSRPETTDALILT-DMAFSNFYSGIKSESEVVLACIM 297
QY      452 IYRRIRPADKLPD-----IFKNVYIDKKFKGDTKKYADFVDDKSVVPSDFHA 499
DB      298 AARNIT--EKEDYAFVAALLDVYKKAIDRS-RGDED----- 334
QY      500 LKSMDEKFAKALEKPAVELSKSVTAARAIQADAMANAAYAEKKRFLFAGLREMY- 558
DB      335 -LEGVYDHFHFRYIMEDDSYRLNPEL---KNLPLDALANAMDSRLQSFYMKINQMLYD 390
QY      559 -----PGRALPSDANTTMRKSTG-STIKGYEPQDGA---YNYHTT-----GK 596
DB      391 RYFNHDDGRRLRLETPQIIFWMRVAMGLALK--EODKYTWAITFYNILSTFRYTPATPTLFS 448

```

```

QY      597 GYLEKDPKSDPEFAVENIDLEFRTKNYGRVANGOLHIAFLS-----NNDITG-GNSG 649
DB      449 GMRHSQSLSCSYLTVQDDLVNITKV-----ISDN-----AMLSNMAGISGNDWTAIRATG 498
QY      650 SPYFDKNGRLIGL 662
DB      499 ALIKGTNGKSGQV 511

RESULT 9
CN16_HAEIN
ID      CN16_HAEIN STANDARD; PRT; 657 AA.
AC      P44764;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      2,3'-cyclic-nucleotide 2'-phosphodiesterase precursor (EC 3.1.4.16).
GN      CPDB OR H10583.
OS      Haemophilus influenzae.
OC      Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC      Haemophilus.
OX      NCBI_TaxID=727;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=RD / KW20 / ATCC 51907;
RX      MEDLINE=95350630; PubMed=7542800;
RA      Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA      Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA      McKenney K., Sultion G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA      Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA      Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA      Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA      Fine L.D., Frichman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA      Gnehm C.U., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA      Venter J.C.;
RT      "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT      Rd.";
RL      Science 269:496-512(1995).
CC      -1- FUNCTION: THIS BIFUNCTIONAL ENZYME CATALYZES TWO CONSECUTIVE
CC      REACTIONS CONVERTING 2',3'-CYCLIC-NUCLEOTIDE TO 3'-NUCLEOTIDE
CC      AND THEN 3'-NUCLEOTIDE TO NUCLEIC ACID AND PHOSPHATE
CC      (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: Nucleoside 2',3'-cyclic phosphate + H(2)O =
CC      nucleoside 3'-phosphate.
CC      -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
CC      -1- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U37240; AAC22242.1; -.
DR      TIGR; H10583; -.
DR      InterPro; IPR002224; 5_nucleotidase.
DR      InterPro; IPR000934; Ser_thr_phosphatase.
DR      Pfam; PF01009; 5_nucleotidase; 1.
DR      Pfam; PF02872; 5_nucleotidase; 1.
DR      PROSITE; PS00785; 5_NUCLEOTIDASE_1; 1.
DR      PROSITE; PS00786; 5_NUCLEOTIDASE_2; 1.
KW      Hydrolyase; Multifunctional enzyme; Periplasmic; Signal;
KW      Complete proteome.
FT      SIGNAL 1 26 POTENTIAL.
FT      CHAIN 27 657 2',3'-CYCLIC-NUCLEOTIDE 2'-
FT      FT PHOSPHODIESTERASE.
SQ      SEQUENCE 657 AA; 72763 MW; 201CAAB15014499 CRC64;

Query Match 3.0%; Score 112.5; DB 1; Length 657;
Best Local Similarity 18.2%; Pred. No. 8.7;

```

Matches 122; Conservative 82; Mismatches 193; Indels 273; Gaps 36;

QY 187 FLIVDVDFKDVMMVAPAPSSVGKFG-----GDTNNMMWRHTGDF 226
 DB 45 FLTDFEYVD-----AP---TDKFGTTRASTLRQARAAYKKSVLVNDGLDIGNDIADY 96
 QY 227 SVFRYAGADNRPA-----EYSKDNKPKYPYFAVNSOGYK 263
 DB 97 QAAOGYKEKSNPAIDCLNMMNYEVCTLGNHFNGLYTLAD--AIKQAKFPIVNSVYK 154
 QY 264 ADDYANTIFPGSTDRYILTSKGVEDR--IENENRPIEVRGKOG-----IMKA 311
 DB 155 A-----GTEEPYFPYVYIEKSSVNDG---KTKKLKIGIGFPPQIMWMDA 200
 QY 312 -MSADQATR--IKYASKYQASANYMKNISGMNGLARLDVIGRKRAEERAFADWIRKNK 368
 DB 201 NIGKVEREDYKTKQKYVE-----MKKG- 226
 QY 369 SAIVYGDVLSL-----EKAYKEGAKAN---REMYTSETLFGTEVY---REAQFAN 414
 DB 227 ---ADIVVALAHNTGSPDEPYOEGAENSAPYLADVPHIDAVIFGHSRLFPNKKEFAKSPN 282
 QY 415 ALATN-----PDANAG-----ILKSLDDKYYKDXLPISLDKRYLPAMDIVRRIPAD 460
 DB 283 ADIVNGTVGIFESMAGYANNISVVDGLTEHKGMWITSGAAVLRPIYDIETKALAK 342
 QY 461 KLPIDEFKNYIDKKFKGDTKKYADFVDEKSVPYSDKFHMLKSMDEKPAKAIKDPAYE 520
 DB 343 NDEPI--TALLKPVHEATIKKYVSQPIGKA---TDMNYSYL-----ALLQDDPTIQ 387
 QY 521 LSKS-----VIAARAIOADAMAN---AYA-IEKGRLEF-FAGLR 555
 DB 388 IYNOAKAVKVERVAPSIAMAGLPILISAGAPKAGRKNDPTGYEVNKGKILFRNAADL 447
 QY 556 EMVPG-----BALPSPANNTMMSYSGISGYEPQ-----DGAMY 589
 DB 448 YXPTLVVYKATGEOLKLEKLECSAGMEKOIDTSDKPSQLDMEGFRFYNDYIDGVNY 507
 QY 590 NYHTT-----GKGVLEKODPKSPDEFAVOENIIDLFFTKNY 624
 DB 508 EYDLTFPARYDGBCKLINPESHRYVNLTYQKPY---DPKA-EF-----LATNNY 554
 QY 625 GRARAE---NGOLHIAFLFNND-----ITGNSSS--PVFPKNRGLIGLADGWY 668
 DB 555 RAYGNKFPCTGKHIVYASPDERSQIADYIKATSEKESVNPADKNRVPIT--GND 612
 QY 669 EAMSGDIEFE 678
 DB 613 KL---DVREF 619

RESULT 10
 EST2_CABEL
 ID EST2_CABEL STANDARD; PRT; 556 AA.

AC 007085; 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE Esterase CM06B1 (EC 3.1.1.1).
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodermidae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN (1)
 RP SEQUENCE FROM N. A.
 RX MEDLINE=94033620; PubMed=8219278;
 RA Fedon Y., Cousin X., Toulant J.-P., Thierry-Mieg D., Arpaegus M.;
 RT "cDNA sequence, gene structure, and cholinesterase-like domains of an
 RT esterase from *Caenorhabditis elegans* mapped to chromosome V.";
 RL DNA Seq. 3:347-356(1993).
 CC -I- CATALYTIC ACTIVITY: A carboxylic ester + H(2)O = an alcohol + a
 CC carboxylic anion.
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC. ATTACHED TO THE INNER SIDE OF

CC THE MEMBRANE BY A LIPID ANCHOR.
 CC -I- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X66104; CAA46899.1; -.
 CC DR PIR: S21556; S21556.
 CC DR HSP: P37967; I0E3.
 CC DR InterPro: IPR002018; Carboxylesterase-B.
 CC DR InterPro: IPR000379; Est_lip_thioest_acsite.
 CC DR Pfam: PF00135; Coesterase; 1.
 CC DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 CC DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; FALSE NEG.
 CC DR Hydrolase; Serine esterase; Myristate; Membrane.
 CC FT INIT_MET 0 0
 CC FT LIPID 1 1
 CC FT ACT_SITE 207 207
 CC FT ACT_SITE 330 330
 CC FT ACT_SITE 445 445
 CC FT DISULFID 75 97
 CC FT SEQUENCE 556 AA; 62391 MW; 46BEB5B2428DB99B CRC64;
 SQ

Query Match 3.0%; Score 112; DB 1; Length 556;
 Best Local Similarity 20.0%; Pred. No. 7.4; Indels 128; Gaps 20;
 Matches 77; Conservative 54; Mismatches 126;

QY 306 GIKWEAMSDQATRIKAYKQASANYMKNISI---GMNGLARLDVIGRKRAEERAFADW 362
 DB 176 GLM-----DQTLAKWKVGHKHSFEGDPNCVTVFQSGAGSTDLISPSHSRDLFGQF 229
 QY 363 IRRKNGS---AVYGDVLSLEKAYKEGAKANREMYTSETLFGTEVYRFAOF----- 412
 DB 230 IPISTAHDCFAI-----RASENQAKIFRE-----FAEFHFGSGR 264
 QY 413 -ANAL-----ATNPDAHA---GILKSLDDKYYKDXLPISLDKRYLPAMDIVRRIPADKLP 463
 DB 265 DSSALFKWTQEOSPELTSNVKGYKKSISG-FLTFIPNLGDGDFPKLDRKEAP----- 318
 QY 464 DIFKNYIDKKFKGDTKKY-----ADF-----VEDKSVYPYSDKFHAML 501
 DB 319 -----KKQMTGVTEYEGMLASWNPAPFSPADVGLTLPQGIYKDVVSNPDE---I 367
 QY 502 KSMDEKFAKAIKEDPAVELSKSVIAAARAIOADAMANAIAIE-----KGRLEFAGL 554
 DB 368 OKIFEKYEYEGVDKSDDELMMRKLEAL---GDEFFNNGVIOAKNAKHGNEVFFYTF 423
 QY 555 REMYPGRALPSDANFTMRMSYSGISGYEPQDQAMVYHTT-----GKGVLEKODPKSD 608
 DB 424 EYVNDP-----SFGMMDSMPFKAA---VHCTELRYLLGSEVYSKFEPTED 467
 QY 609 FAVOENIIDLFFTKNRYARAE-NGQ 632
 DB 468 RKVMEETTLTF--SNPAKXGNPNGK 490

RESULT 11
 DPOZ_YEAST
 ID DPOZ_YEAST STANDARD; PRT; 1504 AA.

AC P14284; 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE DNA polymerase zeta catalytic subunit (EC 2.7.7.7).
 GN REV3 OR PS01 OR YPL167C OR P2535
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; *Saccharomycotina*; *Saccharomycetes*;
 CC *Saccharomycetales*; *Saccharomycetaceae*; *Saccharomyces*.

```

OX NCB1_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=900080808; PubMed=2676986;
RA Morrison A., Christensen R.B., Alley J., Beck A.K., Bernstine E.G.,
RT Lemont J.F., Lawrence C.W.;
RT "REV3, a Saccharomyces cerevisiae gene whose function is required for
RT induced mutagenesis, is predicted to encode a nonessential DNA
RT polymerase."
RL J. Bacteriol. 171:5659-5667(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RX MEDLINE=97103777; PubMed=8948103;
RA Purnelle B., Coster F., Goffeau A.;
RT "The sequence of 55 kb on the left arm of yeast chromosome XVI
RT identifies a small nuclear RNA, a new putative protein kinase and two
RT new putative regulators."
RL Yeast 12:1483-1492(1996).
CC -1- FUNCTION: REQUIRED FOR INDUCED MUTAGENESIS, IT IS A NONESSENTIAL
CC DNA POLYMERASE. IT MAY FUNCTION IN TRANSLATION SYNTHESIS.
CC TRANSLATION SYNTHESIS IN S.CEREVISIAE MAY USE A SPECIALIZED DNA
CC POLYMERASE THAT IS NOT REQUIRED FOR OTHER DNA REPLICATIVE
CC PROCESSES.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M29683; AAA34968.1; -
DR EMBL: X96770; CAA65554.1; -
DR EMBL: Z75523; CAA97873.1; -
DR PIR: A33602; A33602.
DR SGD: S0006088; REV3.
DR InterPro: IPR002064; DNA_pol_B.
DR Pfam: PF00136; DNA_pol_B: 1.
DR Pfam: PF03104; DNA_pol_B_exo: 2.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLBc: 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B: 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; DNA repair; Nuclear protein; Zinc-finger.
FT ZN_FING 1398 1417 C4-TYPE (POTENTIAL).
FT ZN_FING 1446 1473 C4-TYPE (POTENTIAL).
SQ SEQUENCE 1504 AA; 172956 MW; 34C2C6B64F734F5 CRC64;

Query Match 3.0%; Score 112; DB 1; Length 1504;
Best Local Similarity 17.8%; Pred. No. 29;
Matches 146; Conservative 105; Mismatches 255; Indels 316; Gaps 39;

QY 38 NUDPMRBLGLP-LDSLYSDKRSINAVYTGCGCTGTVSOGGLFTNHHC----- 90
DB 212 NVDRCY--FRSPVLNSILDLDK-----LTINDLQLLDLRFCDKCNV 252
QY 91 -----GVGAIO-----SOSTVDHDLRLD-GPVS-----RTM 115
DB 253 LSRDPRVNGNGLIEILILPOFINKREKLOHRDHHDFLEKGLISDIPKAPYSSADM 312
QY 116 GEELPI--PGLSVKYLKRIKIVKVDKVGQIKGIDEMERL-RKAQEVCOELAKKENADEN 172
DB 313 INELTMORELSLKEKPEPKTRHVSQHOMQSSGEFEAFYKKAQH-----KTSFEDG 365
QY 173 QLCIVEFYSINNEFLVYDFKVRVAFPPSSVKGFGGDTDMWMPRHTGDSVFRVY 232
DB 366 QIPNFENFIKDKFSAINTPYE-----ALPQ-----LWPR----- 396

```

```

QY 233 AGADNRPAYESKDNKPKPYVEAAVSMQYKADDYAMTIGFGSTDRYLSFM---GVEDR 289
DB 397 -----LPOIEINNN-----SMQDKKND-----QVNASFTYEICGVNDE 431
QY 290 IENENPNRIEVRGIRKOGIKWKEAMGADQATRIKYSKQAOSANY-----WK 334
DB 432 NEGKGSNKRKSRYSWLPESIASPKDSFTLLDQTKYHNTINFNSDMCAMQNMASKRLR 491
QY 335 NSIGMNGRLARLDYIGKKRAEERAPADMIRKNGKSANYG-----DYLSSLEKA----- 382
DB 492 SSVGAN-----KTSLSRRK--KKVMAGLRYGKRAFYVGEPPFGYODILNKLEDEGFPKI 545
QY 383 -YKE-----GAKANREMYLSTFL---FGTGV-----VRFAQFAN-A 415
DB 546 DYKDPFSSNPVDLENKPRYAAGKRFELISSHVSSTRIPVQGEVYSYINKRPFMFSSWK 605
QY 416 LATNPDAHAGILKSLDDKYKDYPLDRKVLPAMLDIYRRRIPADKLPRDIKNYIDKKFK 475
DB 606 YALRPPTYDAN-----QKWYKVPVSMGKKTESQISM---HTPHSKF--LYKFPASDVSGK 655
QY 476 GDYTK-----YADVYDKSVVPYSKPFHMLKSMKEXFA----- 510
DB 656 QKRKSSVHDSLTHTLLEIHANTRSDKIPDPAIDEVSMITICLEETFPDLIDIAYESGM 715
QY 511 ---KAIEK-----DPAYELSKSVYAAARAIQADAMAN-----A 540
DB 716 IVHRASESTFPYKIQHCINEIPVMEFYSEFMEFEDALDVLDPDILTSGEFIHNSWG 775
QY 541 YAIKRGKRLF-FAGURE-----MYPG-----RALPSDAN 568
DB 776 YIIRCKQIHQFDIVRELARVCKQIKRKLSDTGWYAHSSGIMIGRMHINWRALRSQVN 835
QY 569 FTMMSYSGISIKGYPQDCAWYNYHTGKGVLEKDDPK-----SDEFAVOENIIDLFRTK 622
DB 836 LTO-----YTIESAAFNILHKRLPHSFESFLTNMMAKSTELKTVL 878
QY 623 NYGRYAENGOLHIAFLSNDDITGNSGSPVDFKNGRLIGLAF 664
DB 879 NY--WLSRAQINIOILKRQDIYARI-----EQARLIGIDF 912

RESULT 12
ACON_RICPR STANDARD; PRT; 878 AA.
AC Q9ZCF4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aconitate hydratase (EC 4.2.1.3) (Citrate hydro-lyase) (Aconitase).
GN ACONA OR RP799.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCB1_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MADRID E.
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Slicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria".
RL Nature 396:133-140(1998).
CC -1- CATALYTIC ACTIVITY: Citrate = cis-aconitate + H(2)O.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- MISCELLANEOUS: ACONITASE HAS AN ACTIVE (4FE-4S) AND AN INACTIVE
CC (3FE-4S) FORMS. THE ACTIVE (4FE-4S) CLUSTER IS PART OF THE
CC CATALYTIC SITE THAT INTERCONVERTS CITRATE, CIS-ACONITASE, AND
CC ISOCITRATE.
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/TPM ISOMERASE FAMILY.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ235273; CA15225.1; -
DR HSSP: P16276; 7ACN.
DR InterPro: IPR001030; Aconitase.
DR InterPro: IPR000573; Aconitase_C.
DR Pfam: PF00694; Aconitase_1.
DR PRINTS: PR00415; ACONITASE.
DR PRODOM: PD000511; ACONITASE_1.
DR PROSITE: PS01244; ACONITASE_2; 1.
DR PROSITE: PS01244; ACONITASE_2; 1.
KW Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S;
KW Complete proteome.
FT METAL 426 (IRON (IRON-SULFUR CLUSTER)
FT METAL 492 492 (IRON (IRON-SULFUR CLUSTER)
FT METAL 495 495 (IRON (IRON-SULFUR CLUSTER)
FT METAL 495 495 (IRON (IRON-SULFUR CLUSTER)
SQ SEQUENCE 878 AA; 97120 MW; 283539115E6605 CRC64;

Query Match 3.0%; Score 111.5; DB 1; Length 878;
Best Local Similarity 19.4%; Pred. No. 15;
Matches 144; Conservative 96; Mismatches 255; Indels 249; Gaps 37;

QY 9 LTGAALILGASGAKADKGMWLNENLDMRELGLPLDLSYSPDKPSIAAVYI 68
DB 242 VIOVKLKLKAGMATKATDVLTTETLRR-----KKVQKFEVER 280
QY 69 FGGCGTGTIVSDGLFTNHHCYGAIOSSTVDHDLRDGFVSRTMGEELPIGLSVKY 128
DB 281 FGNGLHNLITSDRATL-SNNSPEYGATCGFFPIDQETLK-----YLEVTGREITQ 329
QY 129 LRKIVAVTDKVEGQLGIDEMERLKAQE-----VQQLA--KKENADENOLCTIPEPY 181
DB 330 IKLVEYTA--TEONLWNCEDTQETVELELDISTVYSSLAGRPKRODRVNLNCVANSFO 387
QY 182 SNNEEFLIVYDFKDYRMVAPPSSVQKFG-CD-----TDNMWMPRHGTDFSVFRV 231
DB 388 NELPYALAN--KDIDKTYAANONVTAIGNGVIAAITSCI-----NTGNPSVMIG 437
QY 232 YAGADNRPAEYSKDNKPYKPYFAA-----VSMQGYKADYAMTIGFGSTDRYLS 283
DB 438 AALLAKKALEHGLNKPWVYTSLAPSKVYTEYKLISG--LDKRYDALGF-----NLVG 489
QY 284 WGVEDLLENEN--NPRIEYNGIKGIMKEMASDAQATRIKYASKYASOANYMKNISGMNR 341
DB 490 YGCTTGIGSGSLNPELE-----NTINKNR 514
QY 342 GLRLRVYIGKRAEERAFADWIRKNGSAVYGDVLSLEAYKEGAKANEMTYISSETLE 401
DB 515 LVYASVLSGKRNFEGR-----INPLTKA-----SYLSPIL 545
QY 402 GGTEVVRFAQFA--NALATNPDAHAGILSLDKRYDYLPSLDKRYLPAWLDIVRRRIPA 459
DB 546 ----VVAVALSGTLNIDLTNPICANITY-----LKDIMP--QKEIDVANSINSMTI 594
QY 460 DKLPDLFKNVYDKKFGDTKKYAD-----FVFDKSVY-----PYSDKFAH--MLKSM 504
DB 595 EKADVDFNG-----TKEMRDLOITTGTYNMMDKNSTYINNPPYFDNIGSEHSIKDI 645
QY 505 DKKEK-----PAKALEKD--PAVELSKSVTAARAIOADAMA--NAYALEKKG-----RL 549
DB 646 KSRRIILAIIGDSITTTTHISFAGISIKNSPAKAKLKHHEPDLFNSISGRGHEHYMMRG 705
QY 550 FFAGLR---EMYPGRALPDPANFTM-----RMSYGSIKGYEPD-----G 586

DB 706 TFAIRIKRMKNCNG-----VEGFTINQLSGVQGTIVDFAMDYKAHDIPLVIFAGKEVGS 761
QY 587 AWNYHTTKGVLEKQDPKSDERAVQENILDFRTNNGYGAENGOLHIAFLSNNTITG 646
DB 762 SSRDWMAGKGLL-----GIKAVIAESPERIHSNLMVGMLPLTF-----TGK 805
QY 647 NS-----GSPVFDKNGRLIGLA 663
DB 806 NTRLNKLKDGSEIID-----LIGLS 825

RESULT 13
YKLE.CAEEL STANDARD; PRT; 922 AA.
ID AC 42173;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE Hypothetical 103.8 kDa protein C03C10.6 in chromosome III.
GN C03C10.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2.
RA Gardner A., Berts M.
RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z38112; CAA86235.1; -
DR EMBL: Z35637; CAA86235.1; JOINED.
DR EMBL: Z38112; CAA86235.1; JOINED.
DR WormPeP: C03C10.6; CE02959.
KW Hypothetical protein.
SQ SEQUENCE 922 AA; 103818 MW; DE0888AB97FC87D8 CRC64;

Query Match 3.0%; Score 111.5; DB 1; Length 922;
Best Local Similarity 21.1%; Pred. No. 16;
Matches 130; Conservative 77; Mismatches 189; Indels 219; Gaps 33;

QY 55 YSPDKPSIANAVYIFGGCGTGTIVSD-QGLIFNHHCYGAIOSSTVDHDLRDGFVSR 113
DB 151 YRPDQLPVSEYIV-----RLGIVYVKTQONOV--HCSDAADQASADITWDIEDDDVAR 203
QY 114 TMGEELPIP-GLSKYKLVKIVTKDVEGOL-----KGITDEMERLKAQ 157
DB 204 A--ELIREPSLTV--TTIQEATPAEGGVNHGAVCMQPTIPEEGGLAPLELRKRS 257
QY 158 EVCQELAKENA-----DEN--OLCTIV-----EPYSNNE 185
DB 258 TAVQEVRRKKSALAEVKKSCITLRCGHCNEAIEFMFDETTSLCLIALEFLHREPSMAADP 317
QY 186 YPLIYVDYFKDYRMVAPPSSVQKFGGPDNNMMWPHHTDGFV----- 228
DB 318 LFKILYTV--TRLIDTP-----MYPHSTEMFVPANSRSVAKQLRVSLH 360
QY 229 -----FRVYAGADNRPAEY-----SKDNKPYKPYFAVSMQGYKADYAMTIGF 273
DB 361 HLSTSAICQLPFTKIPRDAFMSVALSLADPELSPYFIQIIMEDLEE-----SW 413
QY 274 PGSTD---RYLTSWGVEDRIENENNPRIEYRGIKGIMK--EAMSADQA--TRIKYAS 324

```

Db 414 PGSVKLMKMLAEYVEIPTDMYNNPKDLVGLHLEFFFRKRYSAISADNCITPTBARI-- 471
QY 325 KYQASATYMKNSIGMNGRLARDVIGKRAE---ERAFAMIKKNGKSANYGYLSLEK 381
Db 472 -----ENVITVMTVEYKRYOTFSSSKSPVTLVEFAFMLESLSHA---DV--SLES 517
QY 382 -----AYKGAKANREMTYLSSETLFGGTEVVRFAOPANL-ATNPDAHGIKLSLDDX 434
Db 518 LIGVCAKCANALIRENDKOCITRALV--TELMAIKFKVYLHESNYVTIANMT--LQDAG 573
QY 435 KDY-LPSLDKRVLPAMDIVRRIRPADKLPDIFKNVIDKKKGDYKKAQDFVFDKSVVPY 493
Db 574 EDIEVPLDDQFMTYFASEAIR-----PELF-----EVLDFIADLHYI-- 610
QY 494 SDFHMLKMD-----KEKFAKAIKEDPAVELSK-----S 524
Db 611 -----AKLKESNSDALGDLKVKLAELI---AVEMSRNARDCRTVIRIFWLMSPPS 661
QY 525 VIAARAIOADAMAN 539
Db 662 VTQAASAFADSVTN 676

RESULT 14
MOK1_SCHPO STANDARD; PRT; 2410 AA.
ID MOK1_SCHPO
AC Q9USK8; Q9URT5; Q9URK2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell wall alpha-1,3-glucan synthase moki1 (EC 2.4.1.183).
GN MOK1 OR AGS1 OR SPC338.01C OR SPC1747.01.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Katayama S., Dai H., Arellano M., Perez P., Toda T.;
RT "Fission yeast alpha-glucan synthase Moki1 localizes closely with actin
RT and play a role essential for cell morphogenesis and protein kinase C
RT function."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1210 FROM N.A.
RC STRAIN=972;
RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Murphy L., Harris D.;
RN Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RP [3]
RP SEQUENCE OF 1157-1465 FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Saunders D., Harris D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: UDP-glucose + [alpha-D-glucosyl-(1,3)](N) =
CC UDP + [alpha-D-glucosyl-(1,3)](N+1).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL; AB019183; BAA34054.1; -
CC EMBL; AL023781; CA19332.1; -
CC EMBL; AL049472; CAB39330.1; -
CC InterPro: IPR000461; Alpha_amyase.
CC InterPro: IPR001296; Glycosyltransf_1.
CC Pfam; PF00128; alpha_amyase; 1.
CC Pfam; PF00534; Glycosyltransf_1; 1.

```

```

KW Cell wall; Transferase; Glycosyltransferase.
FT CONFLICT 256 256 F -> Y (IN REF. 2).
SQ SEQUENCE 2410 AA; 272120 MW; 8BD944BAE3A9A3C6 CRC64;

Query Match 3.0%; Score 111.5; DB 1; Length 2410;
Best Local Similarity 17.5%; Pred. No. 58;
Matches 131; Conservative 132; Mismatches 232; Indels 253; Gaps 37;

QY 88 HHCYGAIQSQSVDDHDLADGQVSRIMGEELIPGISVY-----LRTYK 134
Db 1355 HOKFGAV-----GVSNYGRKRWAPYIFWGLKTKG 1387
QY 135 V-----TDK-VEGOLKGTID--EMERLR-KAOEVCOLAKKENADEMOQIVCEPYS 182
Db 1388 LRPNDPTDDEIYDQKVALITDIDPDDEKSKVEKRIAOQWAGLEVENKIDLVFPGRW 1447
QY 183 NNEFYLYVDVFKD-----VRWVFAPP--SVYKFGGDTDNMMWPRH--TGDFSVER-- 230
Db 1448 SQKGIDLIADIASLLESYKQVLICVGPITIDYGFKAERKIDVLQKKYPTRVFSQPKFTQ 1507
QY 231 ----VYAGADNRPAEYKSKDKPKPYFAVSMQ-GYKADDIYAMTIGFPGSTDRYLTSG 285
Db 1508 LPPIYFSGADFAI-----PSRDEPFGLVAVEFGK--ALGIGARVGLGQMPGW 1557
QY 286 VEDRIENENNPRIEVRGKIGIKWKAEMADQATRIKYSKAO----- 328
Db 1558 YS--VESSATPHL-LKQFOAC-QQALSSQRTARARAKSAKORFPVSGMKALELTD 1613
QY 329 ----SANYWKNSTG-----MNRGLARLDVI-----GRKRAEERAFADWIRKNG 367
Db 1614 GCATCSQKRYNRSRNSRSFSLIHESFSRSSEVLPSTSDTNLAKRAE--AEMIMET 1670
QY 368 KSAVYGDVLSLEKAYKEGK-----ANREMTYLSSETLFGGTEVVRFA 410
Db 1671 PPTAEANTGAKLRSLSLGSRGPGHTTEDASDGLDTTQIEESMTAGDSTSGSDISRY- 1729
QY 411 OFANALATNPDAHAGILKSLDDYKDYLPGLDKRVLPAMDIVRRIRPADKLPDIFKNVI 470
Db 1730 ----RAERLNPDSSHSPSEYSTDG----- 1749
QY 471 DKRFKGTFRYADFV-FDKSVPEYSDKFHMLKSMDE---KFAKAIKEDPAVELSKS 524
Db 1750 DYEFDPQRSYYDDLFDDDTTIRNAPSPFQMSFDEHVAVFQSDDDSDP----- 1802
QY 525 VIAARAIOADAM-----ANAYAIKGRKLPFAGLRERMPGALSLDANFTWRMS 574
Db 1803 ----ARVSDSDSVSPPLPPFVAGSNPNARNNNNPFYGNLH-----TSSLSLA-- 1847
QY 575 YGSIKGYEPQDGAWYNYHT-----GKV-----LEKODPKS--DEFAVOENILDLFT 621
Db 1848 -SAMSCKEKDFSLTRVEEFTDEDGALRSFSEKLOKLANKSKDOLCIEQYIMKRSERS 1906
QY 1907 FFFHERRAIKIGLOKPKNLHVNELSHSGTESESLNSGQSYDD---IIMTDESNTYOL 1963
QY 672 SGDIPEPDQRTSVDIRVLEMDKW 699
Db 1964 -GDGDF-----KTHGLKFMFLRTYDW 1985

RESULT 15
SYL_MENJA STANDARD; PRT; 1039 AA.
ID SYL_MENJA
AC Q58357;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Isolucyl-tRNA synthetase (EC 6.1.1.5) (isoleucine--tRNA ligase)
DE (IleRS).
GN ILES OR MJ0947.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 17, 2002, 21:41:56 : Search time 68 Seconds
(without alignments)
1811.359 Million cell updates/sec

Title: US-10-008-355-2
Perfect score: 3719
Sequence: 1 M0MKRLKILGALLIGASG.....LFMIDKMGCCPRILQELKLI 712

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB	ID	Description
1	907.5	24.4	716	16	Q9PC94	Q9PC94 xylella fas
2	146.5	3.9	1368	16	Q99ZM2	Q99ZM2 streptococ
3	133.5	3.6	1104	2	Q06812	Q06812 micrococci
4	126	3.4	953	2	Q48502	Q48502 lactobacilli
5	125	3.4	1440	5	Q9TYU6	Q9TYU6 caenorhabdi
6	124	3.3	1808	10	Q9LOJ8	Q9LOJ8 arbidopsids
7	123.5	3.3	713	2	085051	085051 moraxella c
8	123	3.3	1289	16	Q9X087	Q9X087 thermococ
9	122.5	3.3	742	2	Q9ZB47	Q9ZB47 streptococ
10	122.5	3.3	1571	12	Q92B35	Q92B35 listeria in
11	122.5	3.3	2204	12	Q99F11	Q99F11 porcine tes
12	122	3.3	721	16	Q92BR7	Q92BR7 listeria in
13	122	3.3	743	16	Q9PPY9	Q9PPY9 ureaplasma
14	122	3.3	883	16	Q97S45	Q97S45 streptococ
15	121	3.3	1208	10	Q9LIP5	Q9LIP5 oryza sativ
16	120	3.2	666	17	Q27798	Q27798 methanother

17	118	3.2	698	16	Q9KUZ7	Q9KUZ7 vibrio chol
18	118	3.2	882	16	Q99XS1	Q99XS1 streptococ
19	117.5	3.2	2204	12	Q99F17	Q99F17 porcine tes
20	117	3.1	912	2	006915	006915 lactobacilli
21	116.5	3.1	1273	12	Q9QSK2	Q9QSK2 chilo iride
22	116	3.1	2080	3	Q92215	Q92215 cochlioniou
23	115.5	3.1	1154	16	Q92K97	Q92K97 helicobacte
24	114.5	3.1	561	16	Q9KB24	Q9KB24 bacilli ha
25	114.5	3.1	1198	5	Q9Y059	Q9Y059 caenorhabd
26	114.5	3.1	1306	2	Q47766	Q47766 enterococc
27	114.5	3.1	1530	16	Q9KC46	Q9KC46 bacillus ha
28	114.5	3.1	1758	5	Q22830	Q22830 caenorhabd
29	114	3.1	711	2	Q51284	Q51284 neisseria m
30	114	3.1	901	2	Q9XCV8	Q9XCV8 lactobacilli
31	113.5	3.1	765	16	Q9TFF1	Q9TFF1 neisseria m
32	113.5	3.1	1171	5	000936	000936 toxoplasma
33	113	3.0	632	17	Q97723	Q97723 thermoplas
34	113	3.0	759	16	Q92A91	Q92A91 listeria in
35	112.5	3.0	767	16	Q9JYF6	Q9JYF6 neisseria m
36	112.5	3.0	957	17	Q28565	Q28565 archaeoglob
37	112	3.0	554	5	Q16351	Q16351 caenorhabd
38	112	3.0	804	10	Q9LXB6	Q9LXB6 arbidopsids
39	111.5	3.0	1164	16	Q9ZJN3	Q9ZJN3 helicobacte
40	111.5	3.0	3223	2	Q9RM48	Q9RM48 escherichia
41	111.5	3.0	7659	5	Q9N4M4	Q9N4M4 caenorhabd
42	111	3.0	691	5	Q9GSD1	Q9GSD1 plasmodium
43	111	3.0	972	5	Q9U5P5	Q9U5P5 haemophilus
44	111	3.0	1092	2	Q54099	Q54099 streptococ
45	111	3.0	1713	9	Q94MA1	Q94MA1 lactococcus

ALIGNMENTS

RESULT 1
ID Q9PC94 PRELIMINARY; PRT; 716 AA.
AC Q9PC94;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN XF1887.
GN XF1887.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.U.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Carri C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferto J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Fritume M., Furian L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshaho M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL; AE004008; AAF4693.1; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BDFC CRC64;

Query Match 24.4%; Score 907.5; DB 16; Length 716;
 Best Local Similarity 31.4%; Pred. No. 9.4e-52;
 Matches 239; Conservative 135; Mismatches 291; Indels 95; Gaps 21;

1 MWMKLSILGALLGASGVAKADKGMWLNELNENDRRELGTPLPLDLSYSPDKP 60
 1 MEFNLISVLTATLTVD--THAGEGMWVPOOL-PELAGPIKQGLQSPQLSNLIGD 57
 61 SIANAIVIFGGCGTITVSDGLFTNHCYGAIOSSTVDHDLRDGFVSRTGSEELP 120
 58 PGAVAVSL--GNCPTASVSPGLVITNHCAYGAIOINSTPKNLKEGFNLQDAEVS 115
 121 T-PGLSVTKRIVYVTDKVEGOLGITDEMERLKAQEC--QELAKKENDENQCLIV 177
 116 AGPNARILVLEQITDVTQAKAALAAAGNDPFRRTTALETFSKOEIAKCEE-EGGYRCOF 174
 178 EPEYSNNEFEFLVYDVFVKDVRMVFAPSPSVGKFGGDTDMWMPRHGTGDFSVRYAGADN 237
 175 EEFAGNTYRVKKNLEIKDVLVAPQGSVGFEGGDVDMWMPRHGTGDFSVRYAGADN 234
 238 RPAEYSKDNKPKPYFAAVSNQGYKADYATMTIGFPGSTDRYLTWSGVEDRIENENNR 297
 235 KPASFSEKNIYPRPKHMLKFSQPLGDDGFVAVAGYPRTRNYALV-----AEEFNTHA 288
 298 IEVRIKO-----GIMKEAMSDADATRIKYASKYAQASNYKNKSTGNMRLARLDVIGR 351
 289 WTYPIVIGHEFKMLIALIEAASKNDPIQVKAATLGNNTSKNPDGQLDGRNAINAQ 348
 352 KRAEERAFADWIRKNG---KSAVGYDVLSTLEKAYKEGAKANREMTYLSFTLFGTEYV 407
 349 KOSEETAVLAWIKOQGINHEALAHQTLVDLDEY---KANQDRPIVIG--FNGSGVI 403
 408 RPAQANALAA---TNPDAH-----AGLIKSIDDKYKDYLDPSLDKV----- 445
 404 GYAVNLVYLRLAERTSDAQREAGYQERDLPTEGNLKWME---RRYLEMDRQMOQYWT 460
 446 ----LPMADIVRRRLPA-----DKLPDIEKNVIDKKRGKGTXYADVFYKSVVPSYD 495
 461 EYNKLP-----VKQVAALDVMGLDGIPTLRLDGTKLSSSEER----- 500
 496 KEHAMLKSMDEKFAKALEKDPAVELSKSVIAARAIOADAMANAVALIEKGRLEFAGIR 555
 501 ----LKMFNADRAAFESSQDPAIRYAVAIMPALIEIEONKIRIRGELLKAPITLQALA 555
 556 EYTP--GRALPBDANFTMRMSTGSIKYEPODGMVNTHTTGKYLEKODPKSDEPAVOE 613
 556 DYNKSHGKVVYDANSSILTFGHVKGYSPKDGEVETPTTLQGYMAK--NTEVEPDSER 614
 614 NLLDFRTKNGRYAEN--GQHIAPLSNNDITGNSGSPYEDKNGKRLITGLAFDGNWEM 671
 615 SLINAKKASTANLADQRTGIVPVLSDLDITGNSGSEYLDANIGKLVGLAFDGNWESV 674
 672 SGDIEFEPDLQRTISVDIRYVLEMDKMGQCPRLIQEELKL 711
 675 SSMWEDPVMRTIAVDSRYQWIMTEVAAPAPHLKELNL 714

S1D26

RESULT 2
 0992M2 PRELIMINARY; PRT; 1368 AA.
 AC 0992M2;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHEICAL PROTEIN SPY1046.
 GN SPY1046.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SE370 / ATCC 700294 / SEROTYPE M1;
 RX MEDLINE-21192684; Pubmed-11296296;
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Pirmeaux C., Sezate S., Suvorov A.N., Kanton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
 RT "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*,"
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
 DR EMBL; AE006549; AAK33936.1; -
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 1368 AA; 158440 MW; 07D04F0B5965762F CRC64;

Query Match 3.9%; Score 146.5; DB 16; Length 1368;
 Best Local Similarity 19.5%; Pred. No. 0.8;
 Matches 167; Conservative 105; Mismatches 228; Indels 357; Gaps 43;

35 NOENLDRMRELGTPLPLDLSYSPDKPSIANAVIFGGCGTITVSDGL--IFTNHCY 92
 609 NEENEDILEDVLTITLFE---DREMIERLKTATN---LFDKVKYKOLKRRRYIGW 659
 93 GAI-----OSSTVDHDLVLR-DGVSR-----TMGEELPIGSL----- 125
 660 GRLSKRLINGIRDKSGKTI--LDFLKSDFANRNMQLIHDSLFKEKDIOKAQVSGOD 718
 126 -----VKYLRKIVKYTD--KVEGQ-----LKGITDEM 150
 719 SLHEIHLANLAGSPAIRKKGILQTVKAVDELVKMGHKNRENIYEMARENQTTOKQKNSR 778
 151 ERLRAQECYCEELAK---KENADEN-QLCIVEPEFSNNEFLIVY-----DVEKDVRAVF 201
 779 ERMKRILEGIRKLSQILKEHVENTOL-----QNEKLYLYLVQNGRQVYDQELDI 830
 202 APPSSVGFGGDTDMWMPRHGTGDFSVRYAGADNRPAREYSKDNKPKPYFAAVSNQ 261
 831 -----NRLSDYDVH-----IVQOS 845
 262 YKADYATMTIGFPGSTDRYLTWSGVEDRIENENNRPIEVRGIGKQIMKEAMSAQDARIK 321
 846 FLKDD-----SIDKVLTRSDKNRKSQDNVSEEVYKMKMYKQOLNAKLITQK 896
 322 YASKYAQASNYKNKSTGNMRLARLDVIG---RRAEERAFADWT-----RKNGK--- 368
 897 F-----DNILKAERG---GLSELDKAGFIKRLQVETRLQIKHVAQLILDSMRNRYDEN 946
 369 -----SAVYGDVL-----SLEK 381
 947 DKLIREVAVITLKSILVSDPKDFQFYKREINNYNHAADAYLNAVGTALIKKYPKLES 1006
 382 AYKEG-----AKANREMTYLSFTLFGTEVFAQFANALATNPDAHAGILKS- 429
 1007 EYVYGDYVYVYRKMIAASEQETGATKATYFFYSINIMFKEITLANGEIRKRLIETN 1066
 430 -----LDDKYKDYLPSSLDKVLV-AMLDIVRR-----IP---ADKLPIEFK 467
 1067 GETGEIVMDKGRDF--ATVRYKVLSPQVNIYKTEVGTGFGFSKESILPKRNSDKL----- 1119
 468 NVIDKKFKGDTKKYADVPFDSKSVVPSYSDKFA-----MKKS-----MKKEFA 510
 1120 --IARRKDWDPKRYGG--FDSPYAVSVLVYAVKVEGKSKKLKSVKELLGITIMRSSFE 1175
 511 -----KALEKDPAVELSKSVIAARAIOADAMANAVALIEKGRLEFAGIRMY 558
 1176 KNPIDFLAKGYKEVKKDILIKLPK-----YSLFELENGKRRKLASAGELQ 1221

```

RESULT 3
006812 ID 006812 PRELIMINARY: PRT: 1104 AA.
AC 006812;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALPHA-AMYLASE (EC 3.2.1.1) (1,4-ALPHA-D-GLUCA GLUCANOHYDROLASE) .
OS Micrococcus Sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcaceae; Micrococcaceae; Micrococcus.
OX NCBI_TaxID=1271;
RN [1]
RP SEQUENCE FROM N.A.
RF STRAIN=207;
RX MEDLINE=91115066; PubMed=2276614;
RA Kimura T., Horikoshi K.;
RT "The nucleotide sequence of an alpha-amylase gene from an
RT alkalophilic thermophilic Micrococcus sp. ";
RL FEWS Microbiol. Lett. 59:35-41(1990).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
DR EMBL: X55799; CAA3931.1; -.
DR InterPro: IPR000461: Alpha_ amylase.
DR Pfam: PF00128; alpha-amylase; 1.
DR Hydrolyase: Glycosidase; Carbohydrate metabolism.
SQ SEQUENCE 1104 AA; 121771 MW; E2A0AB6035602B75 CRC64;

```

[illegible]

```

RESULT 4
Q48502 ID Q48502 PRELIMINARY; PRT; 953 AA.
AC Q48502;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE ALPHA-AMYLASE.
GN AMYA.
OS Lactobacillus amylovorus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_Taxid=1604;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CIP 102989;
RA Giraud E., Cuny G.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE OF 1-155 FROM N.A.
RC STRAIN=NRRLB4540;
RX MEDLINE=95077354; PubMed=7986030;
RA Fitzsimons A., Hols P., Jore J., Leer R.J., O'Connell M., Delcour J.;
RT "Development of an amylolytic Lactobacillus plantarum silage strain
  expressing the Lactobacillus amylovorus alpha-amylase gene.";
RL Appl. Environ. Microbiol. 60:3529-3535(1994).
DR EMBL: U62096; AAC45781.1; -.
DR EMBL: X80371; CAA56557.1; -.
DR HSSP: P00691; 1BAG.
DR InterPro: IPR000461; Alpha-amylase.
DR Pfam: PF00128; alpha-amylase; 1.
DR Pfam: PF02806; alpha-amylase; C; 1.
SQ SEQUENCE 953 AA; 104990 MW; F6ABDEDE5619C5F3 CRC64;

Query Match 3.4%; Score 126; DB 2; Length 953;
Best Local Similarity 19.0%; Pred. No. 10;
Matches 115; Conservative 83; Mismatches 246; Indels 162; Gaps 27.

QY 207 VGKEGGDVF--DNMMMPRHNGDSEVFRVYAGADNRPAEYSKDNKPYKRVYFAAVSMGQYKAD 265
DB 95 VGNSSGSKSLNNMYWLYQPTKYSTGNVYLGTEAEFKSMCAAAEYNIIRITVDATLNDTTSQ 154
QY 266 DYAMATIGFPGSD-----RYLTSMGVEDRIE-----NENNPRIEV-----RGI 303
DB 155 YSAISDEIKSISDWTHGNTQISNMGDREYVQNSLLGFYDWMYQNSQVOTYILKNHERLI 214
QY 304 KQGIKKEAMASADQATRIKAYASKY--AQSANYKKNSIGMNGIARLDVIGRKRAEERAFAD 361
DB 215 SDG--ASGGRYAANHIEELPSQDSYSGENFNP----- 246

```

Matches	108;	Conservative	72;	Mismatches	162;	Indels	194;	Gaps	30;
Qy	132	IVK	-VTDKVEGOLKGTIDEMERLKRQEOYCOBLAKKENADENQCTIVERFFYSNNREFLIV	190					
Db	237	LVKGVADRDQGEVVKYIAEIM	-----LSVCGEVYKRESTRK-	271					
Qy	191	YDVFVDKRVKVEAPPSPVSGKFGGDTNNMMPRHTGQSFVRVYAGADNRPABEYKSKDNKPYK	250						
Db	272	-----RELEAFQ-	-----DVKWMTPBQLGITLALKA-AGNDSVEQAKVKE-----	311					
Qy	251	PVYFPAANSMQGIKADDIYAMTIFGPFSTDRYLTSMGVEDRIENENNPRIEVRGIRKQIMKE	310						
Db	312	--FEGOLPADQOK--	-TLTEEFKCKCKAYFFPLMTSELE-----KIKTLKGD--KE	356					
Qy	311	AMSAQOATRIKAYSKYAAOSANYKMSIGMNRGLARLDYIGRKRAEERAFADIRNRKSKSA	370						
Db	357	AAGA-----	-----LVKG--	-VYDROGEVKAIAE-----	378				
Qy	371	VYGQVLSLSEKAYKRGKAGKANREMTYLSLELFGSTGEVYVPAQQA	-----NALATN	419					
Db	379	---KMLITGCEGYKXKSSRRRREIE	-----AEQDFVKMMTPBQLGDYNALKLT-	422					
Qy	420	PDAAHGILKSLDDKKYKD--	-LPSLDRKVL-----PAM--	-LDIVRRIRPA	459				
Db	423	---AGKSEIQQDKKEYFALLPSQOQTLTEEFKQKCGKVYFTPLMTSEELDKI-KTMD	477						
Qy	460	DKLK--DLEFKNVTDKKPKGDTKKVADPFQDKSVNVPYSOK	-----FHMLTKSMQKE	507					
Db	478	DKVAAAGALYKGVADQQ--EGEVKALIEKMLSYGGEVYKDSKRRRELEAFKRFVIMMTPPE	536						
Qy	508	KFA-----KAIEKDPAVELIS-KSVIAAARAIOADMANNAVIAIEKCK-RLFEAGLREMPYG	560						

RESULT 6	
ID	PRELIMINARY; PRT: 1808 AA.
AC 09LOJ8	
DT 01-OCT-2000 (TREMBLrel. 15, Created)	
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE F28G4.7 PROTEIN.	
GN F28G4.7.	
OS Arabidopsis thaliana (mouse-ear cress).	
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;	
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.	
OX NCBI_TaxID=3702;	
LN []	
RP SEQUENCE FROM N.A.	
RA Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,	
RA Alatafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,	
RA Buehler E., Chao Q., Chin C., Choi E., Gonzalez A.,	
RA Hong B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,	
RA Leitz C., Liu A., Liu S., Mukherjee N., Pham P., Sakano H., Shinn P.,	
RA Toitum M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.,	
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.	
DR EMBL; AC007843; AAF97311.1; "	
DR InterPro; IPR003489; Ribosomal_S30.	
DR SEQUENCE 1808 AA; 202730 MW; 9B805C7E2B53EB1 CRC64;	
Query Match 3.3%; Score 124; DB 10; Length 1808;	
Best Local Similarity 18.7%; Pred. No. 39;	
Matches 157; Conservative 113; Mismatches 289; Indels 280; Gaps 38;	
QY 23 KADKCMMLNEL--NQENLD-----RMRELGFILP-----LDLSLX 56	
DB 643 KSNELIPLINDIQRCQTWLDARASKSGARNFVEIAKVRKLLHCFIWDYFSSLS 702	

RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richards D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.,
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of *Thermotoga maritima*.";
RL Nature 399:323-329(1999).
DR EMBL: AE001761; AAD36071.1; .
DR TIGR: TM0992; .
KW Hypothetical protein: Complete proteome.
SEQUENCE 1289 AA; 148491 MW; 8255081AD98ADCAA CRC64;

Query Match 3.3%; Score 123; DB 16; Length 1289;
Best Local Similarity 18.1%; Pred. No. 26;
Matches 125; Conservative 111; Mismatches 256; Indels 200; Gaps 29;

QY 102 DHDYRDGVSTKMGELPIPLSVKYLKTYKVDKVGQLKLTID----- 148
DB 22 DSDALR-YRLMTISE-----LNKSYLKIQNIIVG-YNGLRSGSDFLETLYSNDPLL 72
QY 149 -----EMERLKAQEVCOELAKKENADENQCIPEFYSNNEFL----- 188
DB 73 ARYQKESQAIKRYQARREGASNEIILKQLATQEARLREKQFIALIESNLNSIDYL 132
QY 189 -----IVYDFEKDVRVAFPPSSVGRFGGDTDNMMWRATGDSV-FRYVA 233
DB 133 ASETLKNLSKTANKIIDIIVLKMKEILKPLE-GDGVGYIKNALIQLIDSTYKIFIDYC 191
QY 234 -----GADNPAPYSKDNKRYKPYFAAVSMOGYKADYMTIGFPGSTDRYLT--SMGVE 287
DB 192 KANYGATEKIAQY-----WMKTYF---VEPFESSEKKVL-----DEVLTAKSDEIK 235
QY 288 DRIENENNRIPIVGRIGKOG-----IMKRA--MSADQATR-----IKVASYAQA 330
DB 236 DLKRLRLRLLEIYSKGNALAEKVIETAKKSAESILKNIETPSLIVELFTKYINV 295
QY 331 NY--WKNSIGMNGRL--ARLDVIGKRAE-----ERAFADWIRKNGSAVYGVLSL 379
DB 296 DQIIFNDIASNEVYFIKQIRELVGNDENEINRCYFDRAYFLSKKKAAMINDVGEKT 355
QY 380 EKAYEGAKANEMTYLSTLFGTEVYVFAQFANALATNPAAHAGILSLDK----- 433
DB 356 QGGSTKGQKSEEDVIAPOEVIAKEIY-----LAKDPKLSAVONITRIEEMND 407
QY 434 --YKDYLPSLDR-----KVLPAAMDIVRRIRPADK 461
DB 408 LSLRALPRLNSAKRLIETFGOLENNEITLGEYNEIRIYSVTTGOLNTAKKITANNL 467
QY 462 LPDIFKNVI-----DKRFKGGTKKYADVF-----DKSVVYSDKFAHMLKS 503
DB 468 ESEYHKGISQSEYSEPKLQDLKDVEGNIRQLNDPFYLLKHQELKQROKRFKLSQLVYT 527
QY 504 MKKEFAKAIKEDPAVELSKVIAAARAIQADAMANAIAIEGKRLF--PAGLREMYG 560
DB 528 LDQGGEMKL-----LQDSILA-----KFNQFLKLYQKIGQAVWSAFSFSFSEMMG 574
QY 561 RALPSDANFTMMSYGSIGYEPDQAMYNHTTG-----KGVLEKQPKSDEF-A 610
DB 575 FVINGGKAPMKELNSKSPFLLQNGGWTLENLAIQEDERLTRLILEKVDKLSBELKA 634
QY 611 VOENILDFRTKNYG--RYAENGOLHIAFLSN 640
DB 635 LVESFPRCYTDEDGLVYSYKNAAMLEEFKSS 666

RESULT 9

Q92B47 PRELIMINARY; PRT; 742 AA.

AC Q92B47;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

DE CPA.
GN CPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID:1314;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-CS101;
RA Podbielski A., Weischnik M., Leonard B.A.B., Schmidt K.H.;
RT "Characterization of ure, a global negative regulator gene in group A
RT streptococci.";
RL Mol. Microbiol. 31:0-0(1999).
DR EMBL: U49397; AAC97148.1; .
DR TIGR: 742 AA; 83683 MW; 97A1FF4B4ECB944 CRC64;
SEQUENCE 742 AA; 83683 MW; 97A1FF4B4ECB944 CRC64;

Query Match 3.3%; Score 122.5; DB 2; Length 742;
Best Local Similarity 21.2%; Pred. No. 12;
Matches 133; Conservative 62; Mismatches 212; Indels 221; Gaps 31;

QY 206 SVKFGGDTDNMMWRATGDSVFRVYAGADNRPAEYSKDNKPYKPYFAAVSMOG---Y 262
DB 40 STRAFAGEQS--VNRQSSIDYPMY-GYDSYPRGY-PDYSPLKTYHNLKYNLEGSMDY 95
QY 263 KADYAMTIGPFGSTDRILTSKGVEDRIENEN-----NPRIEVGIGKQIKKEMASADQ 316
DB 96 QAYCFNLTKHFPSKSDSVSQCWYKKLQEGTNEFIKLADKPRIEDQLOONI----- 146
QY 317 ATRIVYASKYASANYWKNSIGMNGRLARLDY-----GRKRAEERAFADWIRK 365
DB 147 -LRILY-----NGYPNRNNGIMKGIPLNALIVTQNAIWTDSAQINPDSFTEAS 198
QY 366 NKSAYVYGDSLSEKAYEGAKANREMYLSETLFGTEVYVFAQFANALATNPDAHAG 425
DB 199 NG---INNOQLGMKRLKELIDPMLGSKYSNKTPEG-----YRLN 236
QY 426 ILKSLDDKRYK---DYLPSLDRK---VLPAMLD---IVRRIRPAD-----KL-- 462
DB 237 VESHDKPFQNLISAIEYVDPDPKGEPPAKTEKTSVIRKRYAGDSKLLGATLKLSQ 296
QY 463 -----PDIEFNVIDKKFKGGTKKYADVFDDK 488
DB 297 IEGSGFQEKDFOSNLGETVELPNCGYLTETSSPDGKIAPRIFRKNKV--FIYOK 354
QY 489 S-----VVPYSKFAHMLKSMKE-----KFAKAIKEDPAVELSKSVIA- 527
DB 355 DGSQVENPKKEVAEPYS--VEAYNDFMDEVLSGFTPYCKFYATNKKSSQVYVCFNAD 412
QY 528 -----AARAIQADAMAN--AYAIKGRLEFPAGIREMYPGRALPSDAN---FTWR 572
DB 413 LHSPPDSYDSGETINPDISTMKEVRYTHAGSDLFKVALR-----PRDTNPEDFLKH 464
QY 573 MSYGSIKGYEPDQAMYNHTTGKGVLEKQDPKSDPEANQ--ENILDFRTKNYRYAEN 630
DB 465 IKKVIKGYK-KKGSYN-----GLTGFQRAAQALAYTTDSADLKTGLTY---NN 513
QY 631 GOLHIAFLSNND-----ITGNGSGS-----PVFDKNGRLIGLAFGNMEA 670
DB 514 GKGYGFESMDKTYLAIVKELITTYAONGSARQLTJNDFPVFNNSDQSLIG----- 564
QY 671 MSGDIEFPDLORTISVDIRYVLEMDK 698
DB 565 -----TECHPD-----DLVDVIRMEDK 581

RESULT 10

Q92B35 PRELIMINARY; PRT; 1571 AA.

AC Q92B35;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE LIN1716 PROTEIN.

```

GN LIN1716.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A:
RX PubMed-11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baguerio F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Chardit A., Chetoui F., Couve E., de Darvar A., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchand E., Durant L., Dussauget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madeno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordisler G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomes of Listeria species.";
RL Science 294:849-852(2001);
DR EMBL; AL596169; CAC96947.1; -.
DR Listlist; LIN01716; -.
KW Complete proteome.
SQ SEQUENCE 1571 AA; 167910 MW; B3B69809D9581F6 CRC64;

Query Match 3.3%; Score 122.5; DB 16; Length 1571;
Best Local Similarity 18.7%; Pred. No. 39;
Matches 138; Conservative 104; Mismatches 308; Indels 187; Gaps 31;

OY 11 GAALLGASVAKAD-----KGMWLLNELNGENLDMRELGTPLDLSYSDKPS 61
DB 559 GGVLLEGKGGKGVKLFSTATAGLKMGKFGKL---ASGVYGLVGLSATLIGMKNN 614
OY 62 IANAVYIFGGCGTGTVDGLIFTNHCGYGA-----IOSQT 100
DB 615 AGEKGTGGFSGSGG--MAGGAIGTMIAFGVTAIGAGTALGKELGYVKEGP 672
OY 101 VHDYLRDGEVST--MGEELPIPLGSLVKYLRKTVK-----YT 136
DB 673 KILDKFKTGMKGLSKTAEBHPILGANININKTIDAAGKIKAGVDTHTVMASKSIYA 732
OY 137 D--KVGGLKGTIDEMERLKAQVCOELAKKENADENQ-----CIYEPPYSNNEY 187
DB 733 DPLKIDASSKGVSKD-----SAKANNEYLNKQKMODSREINVSRRITKELENN--- 784
OY 188 LIYVDVFKDVRVAFPPSSVGFSGDPTDMMWMPRHGTGDSVFVYVAGA---DNRPAYS 243
DB 785 IKTYDKMSD-QLIAATEKKSANKND-----WDKIALGAVSKDPTASAKKHTNNTASTIN 838
OY 244 KDKPKPKPYFAAVSMQGIKADYAMTIGFPGSTDRYLTWSGVEDRIENENNRIEVRGI 303
DB 839 KADIKANNELKLREFFK-----GOEETIRNAEN--RINAL 874
OY 304 KOGIMKE--AMSADOATFRKRYAKYAQASANYMKSIGMNRGLARLD--VYGRKAEERAP 359
DB 875 KQASSEKRLTDEKELO-----RMETTLLEFRSEERST 911
OY 360 ADWIRKNGSAVYDVLSSLKAYKGAANREMTYLSLTFGCTEVRAPQANALAT-- 418
DB 912 EKEVRK-----IEKQRKEAIALATASAKBOKIILGNLENSKKMSAKKAASV 959
OY 419 ---NPOAHGILSLDDKDKDYLPISDRKVLPMPLDIVRRIRPADKLPDIFKAVIDKFK 475
DB 960 VKNSAARDAVSAYEANKYKOTKRIIDK-----RFVYGEISEEYODALKNA--KKK 1011
OY 476 GDRKRYADFEVDSVYPYSDKFAHMLKSMDEKFAKIEKDPAVELSKSVIAAARALOAD 535
DB 1012 NGVYKAEKHDVYREADQQAAGHLREVDME--GETLTK--WEQKKGTVVEFPNSTKDA 1068
OY 536 AMANAVAIIEKGRLLFAGLREMTGRALPSDANFTMMSYSGIKGYEPQDQAMVNYHT-- 593

```

```

DB 1069 ALGKNNELSATVNIIFSGMKDT-ALRVFDSFKTNLKKVINNVITGVN-KVLSFNIKTI 1126
OY 594 -TGKYLEKODPKSDEFVAQENILDLFRKNTNGRYAENCOLHIAFLSNDDTG----GNS 648
DB 1127 VLGNKSGIGSS-----QENSLSAKDKRYTSHYSTQSGLNAMYGSNNASQIMAGEE 1177
OY 649 GSPV-FDKN---CRLIG 661
DB 1178 GEETIAYKNNSAQARILIG 1194

RESULT 11
ID 099FL1 PRELIMINARY; PRT; 2204 AA.
AC 099FL1;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE POLYPROTEIN
OS Porcine teschovirus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC Teschovirus.
OX NCBI_TaxID=118140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VIR 460/88;
RX MEDLINE-21105197; PubMed-11160660;
RA Zell R., Dauber M., Krumholz A., Henke A., Birch-Hirschfeld E.,
RA Stelzner A., Prager D., Wurm R.;
RT "Porcine Teschoviruses Comprise at Least Eleven Distinct Serotypes:
RT Molecular and Evolutionary Aspects.";
RL J. Virol. 75:1620-1631(2001);
DR EMBL; AF296095; AAK12389.1; -.
DR HSSP; Q88590; 1TMF.
DR MEROPS; C03.UPC; -.
DR InterPro; IPR004004; Calic_i_pol_hel.
DR InterPro; IPR001676; Rhv.
DR InterPro; IPR000605; RNA_helicase.
DR InterPro; IPR001205; RNA_pol_psd.
DR Pfam; PF00073; rhv; 2.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR Pfam; PF00910; RNA_helicase; 1.
DR PRINTS; PR00918; CALICVIRUSNS.
SQ SEQUENCE 2204 AA; 247225 MW; FD6B5B5FC6B93BD8 CRC64;

Query Match 3.3%; Score 122.5; DB 12; Length 2204;
Best Local Similarity 19.5%; Pred. No. 66;
Matches 137; Conservative 100; Mismatches 263; Indels 203; Gaps 30;

OY 19 SGYAKADKGMWLLNELNGENLDMRELGTPLDLSYSPDKPSIANAVYIFGGCGTGT 78
DB 572 SEMSEAMQGHVYLYDVGVS-----SDLVIPVSPYDPSPO-----TPYTK 614
OY 79 SDGLIFTNHCGYGAIOSQSTVDHDLRDGEVSRFMGEBELPIGLSVYLRKIVKVTDK 138
DB 615 TDPPYISTN-----MPYVDLPYMSGYVTITQNYTNLASPPSPSQADLTVEYVFG 665
OY 139 VEGQLKGT-----DEMERLKAQVCOELAKKENADENQCI--VEPPYSN----- 183
DB 666 EDVVLKGPSNIFASFOGDTNSPLQAPETGCDRAMTTHASNENPILPVEISGSSRAVFW 725
OY 184 NEYF-----LIYVDVFKD-----VRWVFAPSSV 207
DB 726 DRYFHISTITGVNTRIPQYRLNITSLRRIILROSUHATYRLGSLAIRVAPASPANV 785
OY 208 GKEGDDTMMWMPRHGTGDSVFVYVAGADNRPAEYKDN-----KPYK-PVYFAAVS 258
DB 786 ECDMDGLTVNLVYPPSGMWNININISTSEIQADSDVNDALPSFTWKAQPTVFTCSVP 845
OY 259 MCGYKADYAMTIGFPGSGTDRYLTWSGVEDRIENENNRIEVRGIGKOGIMKEMSADOAT 318
DB 846 YVSF-----TSVLVSYSGFETSYQVPR----- 869

```

```
QY 319 RIKTASKAQSANWKNKNSIGNRGLARDVIGRKRAEERAPADWIRK-----GKSAVY 372
D 870 -----SNOIDPODFGGM-LVLRSSSTTELTJESLWREFKIRLFCPRGIIH 913
QY 373 GDVLSLEKAYKEGAKNARE-----MTYLSLETFGCTEVVFAQANMALNPNHAIL 427
D 914 DS--AAFSKGHPEDSSGSNERLSATMTLS--FGGPGATNFS-----EL 953
QY 428 KSLDDKYDYPLSDRKALPAMLIDVRRIRPA-DKLPDIFKNVIDKFKGDTKKYADFEV 486
D 954 KQAGVEENPGPSLS-KLYQLKD-----PAVDALCEAYDEL--KKRKEGATNLDSDVS 1004
QY 487 DKSVVPYIDK-----HAML--KSMDEKFEKALEKDPAVELSKSVIA--AARAIQADAMA 538
D 1005 GDSENPWINKFKLYGYAILAMKSLHDPMTAAVCFIIGSDITAFVYSKLAKHLKFAKT 1064
QY 539 NAYAIK-----GKRLFFAGLREMYGRALP--SDANFTRMYSGSIKGYEPD----- 585
D 1065 DPPPKPRACKKEGCGCGGNHNYPDELNPSESGFSRFRKGHPQG--PMODVSSLINIL 1123
QY 586 -GAWNYHTTGKGVLEKODPKSDEFAVOENILDFRTKNRGY 627
D 1124 KGAEMVYHOFEKVYKMLKTWRTAREYVSSDFLQA-QMADYHY 1165

RESULT 12
Q92BR7 PRELIMINARY: PRT: 721 AA.
AC Q92BR7
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE LIN1477.
GN LIN1477.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BAguero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihl H., Garcia-del Portillo F., Garrido P.,
RA Gattler L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kunst F., Kurapkat G.,
RA Madueno E., Maitouram A., Mata Vicente J., Ng E., Nedjari H.,
RA Nodules G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001);
DR EMBL: AL596168; CAC96708.1; -
DR ListList; LIN01477; -
KW Complete proteome.
SQ SEQUENCE 721 AA; 79910 MW; D0A0D8167A6837E CRC64;
```

```
QY 243 STA-KEGLPKDAEYLSQGYSRNDRGKSYLEAOYESVLGSKSQSESVLDSKGNIIET 301
D 430 LD---DKYKDYPLSDRKALPAMLIDVRRIRPA-DKLPDIFKNVIDKFKGDTKKYA-DF 484
QY 302 VSKTEGSKGDLVLSVDEFAVEDILRNK-----GQKQYAGSD 343
D 485 VEDRSV---PYSDKFAHL--KSMDEKFEKALEKDPAVELSKSVIAARAIOAD--- 535
D 344 LEFRAVYAMPYSGEVLALGQKLNKGEF-----DDSLGTFTTAVAMGSAVKGSTIL 398
QY 536 ---AMAN-----ATAIEKRLFF---AGIREMYGRALPSDN---FTYRM 573
D 399 GGINDGAISSNKTVEFTDPIALKGTGPKSSWENRKGAGNRPLDPGALIESSNSYMYQVAM 458
QY 574 SYGSIKGYEPQ-----DGAWNYHTTGKGVLE-----EKODPKSDEFAVOENIL 616
D 459 KMGAK-YVNGPLRAPLSTDDMRITYNQGFGVKGIDLPGEQVYKGGDQITGK-IL 516
QY 617 DLFRKNRYAENGQHLIAELSNNDITGNSGSPVDPK-----NGRLGLAFDGNME 669
D 517 DF---ATGQDYSTPLQMAQYVSTIANGSRIAPSMVKELRNSTNDSDVGLTATAN-- 570
QY 670 AMSGDIEFPDLQRTISV---DIRYV 692
D 571 -----EPKVLNKIGVSDNDIKTV 588

RESULT 13
Q9PPY9 PRELIMINARY: PRT: 743 AA.
AC Q9PPY9
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE DNALHELICASE II.
GN UVRD OR U0501.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL: AE002148; AAF30913.1; -
DR HSSP; P56255; 1PJR.
DR InterPro; IPR000212; UVRD-helicase.
DR Pfam; PF00580; UVRD-helicase; 1.
KW Complete proteome.
SQ SEQUENCE 743 AA; 86740 MW; 9D412574673071E7 CRC64;
```

Query Match 3.3%; Score 122; DB 16; Length 743;
Best Local Similarity 17.7%; Pred. No. 13;
Matches 143; Conservative 129; Mismatches 201; Indels 256; Gaps 40;

```
QY 10 LGAALLGASGVAKAD---KGMWLNELNOENLDRMRLEGFTL--PLDSLVSFDRPSI 62
D 22 LGPOLVIAGACTGTSVTLRIAYLTIE--KNIHPSRLGFTTNKAADDEM---KERV 74
QY 63 ANAYVI-----FGGCGTGITVSDGILFTNNHHCIGAISOVSVDHYLRDGFVSKTM 115
D 75 GKTIGVSIPLYSTFRHSMC--VKIILOODIHYLNYH---NNIKIIDTDQEVY----LKEI 124
QY 116 GEELPIPELSVYKRYKIVDKVEGOLKITT-----EMERLRKAQEV----- 159
D 125 FQDLNTEKKS-QVITKIITISKYKKNKFFDQNDMLNEKNHLYLELVLDLNDQRLVDIYKI 183
```



```

QY 160 -COELAKKENADENOLC-----IVEPYSNNEYFLIVDFKVR----- 198
DB 184 YCGRCKLVNVDFFDLINLHKLFIEPEVLEKQMKFDIIV--DEFQOTNKIQYDLIS 241
QY 199 -----MFAPPSSVGKFGGDTDNMMKPHRTGDFSVFRVYAGADNPAEYSKDN-K 247
DB 242 LLATKHQNLFEVVDPPQMIYSFRG-AEOWIINNFSONFK-----NFKVITLKTNYR 291
QY 248 PYKPYFAVSMQGYKADYAMTIGFPGSDRTLTSMGVDRITENENNPRIEVRG---IK 304
DB 292 STQPIILNTANRLIDANNNN-----KNULTAFT-----NDNNLPYLRGQNPID 336
QY 305 QGIV-----KEAMSDQATRIKAYASKYASANYWKNISGMNGLARLDVIGRKRAE 355
DB 337 EAMWIRAKIRLELEGPAQIAVLRSMNHSYTIQ-----SMMESIYITLIGSKFY 391
QY 356 ERA-FADWIRKKNKSAVGYDVLSSLEKAYEGKANKREMTYL-----SETLFGTEYVR 409
DB 392 ERAEIKDMT-----AYLKVVNDLD-----ELSEFLRIINTPRRAIGPTFEHV 433
QY 410 ACPA-----NALATNPDAH-----AGILKSLDDKYYKDYLPSLDRKYLPA 448
DB 434 KHAIAINNNEFLFALAEVENHNIINNTOKKNIINFNVLKEIPEDEMDL-----KIHE 486
QY 449 MLDIVERRI-----PADKLPDIFKNVIDKKFGDTKKYADVPFKSVVPSDKFHAML 501
DB 487 ILELIYKKVVEAYVLENEKAEKIDNVE--LKRAKMYVDRHDPDTINDYINSLALYL 544
QY 502 KSMDEX-----FAAIEKDP--AVELSKSVIAAARAIQADAMANAIAIEKGRKL 549
DB 545 NKDGKSKENVLLMTVHNSKGLYEYENFVAGWMEGLIPSDRAINDPIK--GVEEERRI 601
QY 550 FFAGL-----REMYGRALPSDANTMMSYSGISIGYEPDQAWY 589
DB 602 AYALIRAKKNLXISACCDPLARQVFSRFINELGFNNLKIINSFNNKPED----- 637
QY 590 NYHTGKGVLEKODPKS--DEFAVOENIIDLFRTKNYGRAENGOLHIAELSNNDITGYN 647
DB 658 ---MPLKSFILQOEERSWFSKOKKEVEEDNFIYO-----TMKND----- 693
QY 648 SGSPVFDKNGRLGLAF--DGNMEAMSGDI 675
DB 694 ----FEIGERIVHTSFGDVIITIGDGI 717

RESULT 14
O97S45 PRELIMINARY: PRT: 883 AA.
AC Q97S45;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE VALYI-TRNA SYNTHETASE.
GN SP0568.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.R.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RL pneumoniae.";
Science 293:498-506(2001).

```

```

DR EMBL: AE007366; AA74724.1; .
DR TIGR: SP0568; .
DR InterPro: IPR002300; trna-synt_1a.
DR InterPro: IPR001412; trna-synt_1.
DR InterPro: IPR002301; trna-synt_1le.
DR InterPro: IPR002303; trna-synt_val.
DR Pfam: PF00133; trna-synt_1; 1.
DR PRINTS: PR00966; TRNASYNTHVAL.
DR PROSITE: PS00176; AA-TRNA_LIGASE_1; 1.
KW aminoacyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 883 AA; 100775 MW; F5009144F6A82597 CRC64;

Query Match 3.3%; Score 122; DB 16; Length 883;
Best local Similarity 20.0%; Pred. No. 17;
Matches 102; Conservative 66; Mismatches 167; Indels 176; Gaps 24;

QY 192 DVF-----KDVNFAPSSVGKFGGDTDNMMKPHRTGDFSVFRVYAGADNPAEYS 243
DB 26 DVFKPSGDQKAKPYSIYIIPPNTGKL-----HLG-----HA 57
QY 244 KDKKPKPYFAVSMQGYKADYAMTIGFPGSDRTLT--TSMGVDRITENENNPRIEVRG 302
DB 58 WDT--TLQDIIIRKRMQGF-----TLWLPGRMDHAGIATQAKVEERLREGITRTDGLR 110
QY 303 ----IKGILWKEAMSDQATRIKYA--SKYASANYWKNISGMNGLARLDVIGRKRAEER 357
DB 111 ESFLTWYWEKMD-----EYATTIKEQMGKMGSLVDYSERERTLDEGIS-----KAVR 158
QY 358 APADWIRK-----NGKSAVGD-----VLSSLEKAYK-----EGAKANKREMT 394
DB 159 VFVNLKKKGVIYGEFIIINMDPAARTALSDIEVIHDKVEGAFYHMYMLEDGSGRALEVAR 218
QY 395 YLSEITLFGGEVVRPQAFALATNPDAHNLKSLDDKXKDYLPs-----LDRKYLPA 449
DB 219 TRPEYMGDV-----AVAVNPE-----DPRKIDLGKAVIILPANKLPIY 259
QY 450 LD-----IVRRIRPADKLPDIFKNVIDKKFGDTKKYADVPFKSVVPSDKFHAML 501
DB 260 GDEHADBELGTGVYKITPAIDPNDPLVGQRHNLRYQVNMMDQGTMLAEFGSM--DR 316
QY 497 FHAMLSMDKEKFAKAIKDPAVELSKSVIAAAR-----IQAD 355
DB 317 FEAR-----KAVYAKLEEIGALVKIEKRVHSGVSGTGVVPEPRLSQWFVXMDQAKN 371
QY 536 AMANAYAIKGRKRLFRAGLREMTPGRALPS-----DANTMRSYSGISIGYEPDQ 586
DB 372 AIANQDTEK-----VEYPPRPNDFTFLQWMENVHDWISROLWGH-----QIP 416
QY 587 AMYNYHTTGKGVLEKODPKSDEFAVOENIID 617
DB 417 AMYN--ADGEMVYGEAEPEGDGTQDDEVD 445

RESULT 15
O9LIP5 PRELIMINARY: PRT: 1208 AA.
AC O9LIP5;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE PUTATIVE PHOSPHATIDYLINOSITOL 4-KINASE.
GN DUPR11.7.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20181732; PubMed=10715324;
RA Tarchini R., Biddle P., Wineland R., Tingey S., Rafalski A.;
RT "The complete sequence of 340 kb of DNA around the rice adh1-adh2
RL region reveals interrupted colinearity with maize chromosome 4.";

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 17, 2002, 22:41:11 : Search time 1724 Seconds
(without alignments)
5574.149 Million cell updates/sec

Title: US-10-008-355-2
Perfect score: 3719
Sequence: 1 MQMRLKSILLGALLLGASG.....LFMDKMGCCPRLIGELKLI 712

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus_p2n.model -DEV=xlh
-Q/cgn2.1/USPTO_spool/US10008355/runat_16102002_121153_27707/app-query.fasta.1.903
-DB=EST -QFMT=fastap -SUFFIX=st -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNIT=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human4.0.cdi -LIST=45
-DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355 -ECGN_1_1_1033 -etunat_16102002_121153_27707 -NCPu=6 -ICPu=3
-NO_XLPHY -NO_MAP -LARGEUDERY -NEG.SCORES=0 -WAIT -IONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estlin.*
4: em_estmu.*
5: em_estrov.*
6: em_estrpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss_hum.*
14: em_gss_inv.*
15: em_gss_pin.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	418.5	11.3	591	12	BH392840
C 2	397	10.7	500	12	BH386991

C	3	397	10.7	537	12	BH388186	BH388186 AG-ND-124
C <td>4</td> <td>385</td> <td>10.4</td> <td>418</td> <td>12</td> <td>BH375201</td> <td>BH375201 AG-ND-173</td>	4	385	10.4	418	12	BH375201	BH375201 AG-ND-173
C <td>5</td> <td>315</td> <td>8.5</td> <td>697</td> <td>12</td> <td>BH375688</td> <td>BH375688 AG-ND-120</td>	5	315	8.5	697	12	BH375688	BH375688 AG-ND-120
C <td>6</td> <td>314</td> <td>8.4</td> <td>426</td> <td>12</td> <td>BH400391</td> <td>BH400391 AG-ND-147</td>	6	314	8.4	426	12	BH400391	BH400391 AG-ND-147
C <td>7</td> <td>239.5</td> <td>6.4</td> <td>398</td> <td>9</td> <td>AW774807</td> <td>AW774807 EST333958</td>	7	239.5	6.4	398	9	AW774807	AW774807 EST333958
C <td>8</td> <td>187.5</td> <td>5.0</td> <td>721</td> <td>12</td> <td>BH391233</td> <td>BH391233 AG-ND-141</td>	8	187.5	5.0	721	12	BH391233	BH391233 AG-ND-141
C <td>9</td> <td>184</td> <td>4.9</td> <td>706</td> <td>12</td> <td>BH394089</td> <td>BH394089 AG-ND-150</td>	9	184	4.9	706	12	BH394089	BH394089 AG-ND-150
C <td>10</td> <td>177</td> <td>4.8</td> <td>555</td> <td>12</td> <td>BH400866</td> <td>BH400866 AG-ND-125</td>	10	177	4.8	555	12	BH400866	BH400866 AG-ND-125
C <td>11</td> <td>177</td> <td>4.8</td> <td>769</td> <td>12</td> <td>BH381877</td> <td>BH381877 AG-ND-155</td>	11	177	4.8	769	12	BH381877	BH381877 AG-ND-155
C <td>12</td> <td>160.5</td> <td>4.3</td> <td>1005</td> <td>12</td> <td>BH164253</td> <td>BH164253 ENTTEG637F</td>	12	160.5	4.3	1005	12	BH164253	BH164253 ENTTEG637F
C <td>13</td> <td>138</td> <td>3.7</td> <td>262</td> <td>12</td> <td>BH391877</td> <td>BH391877 AG-ND-141</td>	13	138	3.7	262	12	BH391877	BH391877 AG-ND-141
C <td>14</td> <td>136</td> <td>3.7</td> <td>822</td> <td>12</td> <td>BH371846</td> <td>BH371846 AG-ND-162</td>	14	136	3.7	822	12	BH371846	BH371846 AG-ND-162
C <td>15</td> <td>107.5</td> <td>2.9</td> <td>2349</td> <td>11</td> <td>AK012962</td> <td>AK012962 Mus muscu</td>	15	107.5	2.9	2349	11	AK012962	AK012962 Mus muscu
C <td>16</td> <td>107.5</td> <td>2.9</td> <td>3112</td> <td>11</td> <td>AK004749</td> <td>AK004749 Mus muscu</td>	16	107.5	2.9	3112	11	AK004749	AK004749 Mus muscu
C <td>17</td> <td>103</td> <td>2.8</td> <td>868</td> <td>9</td> <td>AM057494</td> <td>AM057494 ca04.f04.x</td>	17	103	2.8	868	9	AM057494	AM057494 ca04.f04.x
C <td>18</td> <td>102</td> <td>2.7</td> <td>663</td> <td>9</td> <td>AM057482</td> <td>AM057482 ca04.c04.x</td>	18	102	2.7	663	9	AM057482	AM057482 ca04.c04.x
C <td>19</td> <td>102</td> <td>2.7</td> <td>1095</td> <td>9</td> <td>AU205006</td> <td>AU205006 AU205006</td>	19	102	2.7	1095	9	AU205006	AU205006 AU205006
C <td>20</td> <td>102</td> <td>2.7</td> <td>1095</td> <td>11</td> <td>BC015158</td> <td>BC015158 Homo sapi</td>	20	102	2.7	1095	11	BC015158	BC015158 Homo sapi
C <td>21</td> <td>101.5</td> <td>2.7</td> <td>1012</td> <td>9</td> <td>AL545921</td> <td>AL545921 AL545921</td>	21	101.5	2.7	1012	9	AL545921	AL545921 AL545921
C <td>22</td> <td>101.5</td> <td>2.7</td> <td>3340</td> <td>11</td> <td>AK004800</td> <td>AK004800 Mus muscu</td>	22	101.5	2.7	3340	11	AK004800	AK004800 Mus muscu
C <td>23</td> <td>101.5</td> <td>2.7</td> <td>3369</td> <td>11</td> <td>AK011711</td> <td>AK011711 Mus muscu</td>	23	101.5	2.7	3369	11	AK011711	AK011711 Mus muscu
C <td>24</td> <td>100.5</td> <td>2.7</td> <td>651</td> <td>10</td> <td>BE266416</td> <td>BE266416 601193163</td>	24	100.5	2.7	651	10	BE266416	BE266416 601193163
C <td>25</td> <td>100</td> <td>2.7</td> <td>937</td> <td>10</td> <td>BE105351</td> <td>BE105351 601822064</td>	25	100	2.7	937	10	BE105351	BE105351 601822064
C <td>26</td> <td>98.5</td> <td>2.6</td> <td>1131</td> <td>12</td> <td>CNS06LE4</td> <td>AL404066 T3 end of</td>	26	98.5	2.6	1131	12	CNS06LE4	AL404066 T3 end of
C <td>27</td> <td>98</td> <td>2.6</td> <td>667</td> <td>10</td> <td>B1391728</td> <td>B1391728 ppp1n.pk0</td>	27	98	2.6	667	10	B1391728	B1391728 ppp1n.pk0
C <td>28</td> <td>98</td> <td>2.6</td> <td>1087</td> <td>10</td> <td>BE799042</td> <td>BE799042 601583645</td>	28	98	2.6	1087	10	BE799042	BE799042 601583645
C <td>29</td> <td>97.5</td> <td>2.6</td> <td>920</td> <td>9</td> <td>AL545962</td> <td>AL545962 AL545962</td>	29	97.5	2.6	920	9	AL545962	AL545962 AL545962
C <td>30</td> <td>97.5</td> <td>2.6</td> <td>957</td> <td>12</td> <td>A2680449</td> <td>A2680449 ENTHN69TR</td>	30	97.5	2.6	957	12	A2680449	A2680449 ENTHN69TR
C <td>31</td> <td>97.5</td> <td>2.6</td> <td>2836</td> <td>11</td> <td>AK004712</td> <td>AK004712 Mus muscu</td>	31	97.5	2.6	2836	11	AK004712	AK004712 Mus muscu
C <td>32</td> <td>97</td> <td>2.6</td> <td>946</td> <td>10</td> <td>BE685895</td> <td>BE685895 602143193</td>	32	97	2.6	946	10	BE685895	BE685895 602143193
C <td>33</td> <td>96.5</td> <td>2.6</td> <td>2232</td> <td>11</td> <td>BC016514</td> <td>BC016514 Homo sapi</td>	33	96.5	2.6	2232	11	BC016514	BC016514 Homo sapi
C <td>34</td> <td>96</td> <td>2.6</td> <td>3098</td> <td>11</td> <td>AK019547</td> <td>AK019547 Mus muscu</td>	34	96	2.6	3098	11	AK019547	AK019547 Mus muscu
C <td>35</td> <td>95.5</td> <td>2.6</td> <td>509</td> <td>12</td> <td>A2930127</td> <td>A2930127 474.dh250</td>	35	95.5	2.6	509	12	A2930127	A2930127 474.dh250
C <td>36</td> <td>94.5</td> <td>2.5</td> <td>583</td> <td>12</td> <td>BH231647</td> <td>BH231647 110061630</td>	36	94.5	2.5	583	12	BH231647	BH231647 110061630
C <td>37</td> <td>94.5</td> <td>2.5</td> <td>758</td> <td>12</td> <td>BH476716</td> <td>BH476716 BOGS91TR</td>	37	94.5	2.5	758	12	BH476716	BH476716 BOGS91TR
C <td>38</td> <td>94.5</td> <td>2.5</td> <td>772</td> <td>10</td> <td>BJ152144</td> <td>BJ152144 BJ152144</td>	38	94.5	2.5	772	10	BJ152144	BJ152144 BJ152144
C <td>39</td> <td>94.5</td> <td>2.5</td> <td>3200</td> <td>11</td> <td>AK019499</td> <td>AK019499 Mus muscu</td>	39	94.5	2.5	3200	11	AK019499	AK019499 Mus muscu
C <td>40</td> <td>94</td> <td>2.5</td> <td>787</td> <td>9</td> <td>AU140924</td> <td>AU140924 AU140924</td>	40	94	2.5	787	9	AU140924	AU140924 AU140924
C <td>41</td> <td>94</td> <td>2.5</td> <td>846</td> <td>10</td> <td>BE66089</td> <td>BE66089 603295505</td>	41	94	2.5	846	10	BE66089	BE66089 603295505
C <td>42</td> <td>94</td> <td>2.5</td> <td>896</td> <td>10</td> <td>BE583368</td> <td>BE583368 7-8d-xy P</td>	42	94	2.5	896	10	BE583368	BE583368 7-8d-xy P
C <td>43</td> <td>94</td> <td>2.5</td> <td>2426</td> <td>11</td> <td>BC001905</td> <td>BC001905 Homo sapi</td>	43	94	2.5	2426	11	BC001905	BC001905 Homo sapi
C <td>44</td> <td>93.5</td> <td>2.5</td> <td>843</td> <td>12</td> <td>A2687053</td> <td>A2687053 ENTHO85TR</td>	44	93.5	2.5	843	12	A2687053	A2687053 ENTHO85TR
C <td>45</td> <td>93.5</td> <td>2.5</td> <td>1132</td> <td>10</td> <td>BM452650</td> <td>BM452650 AGENCOURT</td>	45	93.5	2.5	1132	10	BM452650	BM452650 AGENCOURT

ALIGNMENTS

RESULT 1
BH392840/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 591)
AUTHORS
TITLE
JOURNAL
COMMENT

BH392840 591 bp DNA linear GSS 11-DEC-2001
AG-ND-162P14.TF ND-7TM Anopheles gambiae genomic clone AG-ND-162P14
, DNA sequence.
BH392840
BH392840.1 GI:17338981
GSS.
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)
Other GSSs: AG-ND-162P14.TR
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org
This clone is from an A. gambiae BAC library (ND-7TM) provided by

F. H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

FEATURES

```
source      1. .591
```

```

/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-MD-162P14"
/clone_1fb="ND-TAM"
/mote="Vector: PEBACBAC1; Site:1; HindIII"
BASE COUNT
ORIGIN
170 a 122 c 105 g 194 t

```

Alignment Scores:

Pred. No.:	1.68e-36	Length:	591
Score:	418.50	Matches:	82
Percent Similarity:	64.58%	Conservative:	42
Best Local Similarity:	42.71%	Mismatches:	63
Query Match:	11.25%	Indels:	6
DB:	12	Gaps:	3

US-10-008-355-2 (1-712) x BH392840 (1-591)

Qy 105 TyrLeuNArgAspGlyPheValSerArgThrMetGlyGluLeuProIleProGlyLeu 124
::: ||| :::: ||||| :::
Db 590 CACCTACCCTATGGTCTCTGGGCAACAGATAATGGTGAGAAATTAAITCTAAAGCTTA 511

Qy 125 SerVallystYrLeuargLysIleVallysvaItmrAsplysvaGluclyglntLeuLys 144
 |||||::: ||:::
 Db 530 TCAGTAGATTCCTTGTAAAGATGGATGATGCGACAAAACGATACCTTCTAAACGAAAC 471

Qy 145 Gly-----Ilethraspglumetglaargleuarqlylsalaaglncluvalcysglnclu 162

Db 470 AATGATATGAGCCGCTGATCACAGAGAAAGCTATTATCGAT-GCTGAATC-----AAA 418

Qy 163 LeuAlaLysLysGluAsnAlaAspGluAsnGln---LeuCysIleValAlaGlnProPheTyr 181

Db 417 GCTATCCAGCTCTGAAAACTCTGAGAAATGGAATAACACTGTGTGTGCTGTAAGATTCTTC 358

QY 182 Serasna^{ngl}uTyrrPheleuilevalTyrraspvalPhe^{lys}aspvalargmetvalPhe 2010

Db 357 AAGGAATGAGTTTACTACTTTGTA^{TTT}CCAGATTTCAAGATGTTGCTGTGAGGT 2988

QY 202 ALAProBrosersValGlyLysPheClGlyAspThrAspAsnTrpMetTrpIArg 222

DB 297 ACTCGCGCTCACTATTGGTAATAATAGGTGTGTACTGTAACCTGGGAATGCCAAGA 238

QY 222 HistThrGlyAspPheSerValPheArgValTyrAlaGlyAlaAlaAspAsnArgProAlaGlu 241
|||||
Db 237 CACACAGAGAGACTTCTCTGTTTCTCCGTCGTTATGCTGATAAATAATGCAATCTCTGTA 178
|||||

QY 242 TyrSerLysAspAsnLysProTyrLysProValTyrPheAlaIaValSerMetGingly 261

262 TyrLysAlaAspAspTyrAlaMetThrIleGlyPheProGlySerThrAspArgTyrLeu 2811

Qy 282 ThrSerTrpGlyValGluAspArgIleGluAsnGlu 293

RESULT 2
BH386991/c

LOCUS BH386991 500 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-148L11.TR ND-TAM Anopheles gambiae genomic clone AG-ND-148L11
ACCESSION BH386991
VERSION BH386991
KEYWORDS BH386991.1 GI:1733133
SOURCE GSS.
ORGANISM African malaria mosquito.
Anopheles gambiae

ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 500)
Sheely,J., Malek,J.,
Direct Submission of BAC end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSS: AG-ND-148111.TF

Euarysta, Metazoa, Arthropoda, Hymenoptera, Hymenoptera, Insecta,
Pterygota, Neoptera, Endopterygota, Diptera, Nematocera, Culicoidea
, Anophelinae.

1 (bases 1 to 500)
Sheely,J., Malek,J.,
Direct Submission of BAC end sequences from Anopheles gambiae
Unpublished (2001)
Other_GSSS: AG-ND-148111.TF

Contact: Brendan J. Loitkus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0508

This clone is from an *A. gambiae* BAC library (20) kindly provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from *A. gambiae* PST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES

```
source      1. .500
```

```

/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-148.11"
/clone_1lb="ND-TAM"
/notes="vector: pECBAC1; Site_1: HindIII"
BASE COUNT      143 a      105 c      90 g      162 t

```

ORIGIN

Alignment Scores:	
Pred. No.:	3,37e-34
Score:	397.00
Percent Similarity:	65.00%
Best Local Similarity:	44.38%
Query Match:	10.67%
DB:	12
Length:	500
Matches:	71
Conservative:	33
Mismatches:	52
Indels:	10
Gaps:	2

US-10-008-355-2 (1-712) x BH386991 (1-500)

Qy 134 LysvalthraspLysvalGluGlyGlnLeuLysGlyIlethraspGluMetGluArgLeu 153
:::||||| :::: ||::: :::
Db 484 CGTATCTCTTAACTGACACATGATATGACC-----GCTGATCAGAGAAACCTATT 431

QY	154	ArgLysAlaGlnGluValCysGlnGluIleuAlaLysLysGluAsnAlaAspGluAsnGln	173
		:	
Db	430	ATCGAGCTGMAATCAAGCATATCCAGCTGTAAACTCTGACAATGCA-----AAATAC	3777

QY 174 LeuGysIleValGluProPheTyrSerAsnGluTyrPheLeuIleValTyrAspVal 193

DB 376 ACTGATAGTGGTGAAGGATTTCTTCAAGAGAAATGATGTTTCTACTCTTTGATTCCAGGAT 317

QY 194 PhelyaspvalargmetvalphealaproproserServalglylyspheglyclasp 2133
|||||
DB 316 TTTCAACATGTTCTGTCGTAGCTACCTCCGCTATCTATTGGTAAATATGATGGTGGAT 257Z
|||||

QY 214 ThrAspAsnTrpMetTrpProAlaGlnHisThrGlyAspPheSerValPheArgValTyrAla 233

Db 148 AAGATCGATTGTCATATTCAATGGGGTGTACTGCAGAAATCGTTCAACAGCGGT 207
 Qy 83 LeuilePheThAsnHisHisCysGlyTyrGlyAlaIleGlnSerGlnSerThrValAsp 102
 Db 208 TTGGTCTTACCAATACACACTGCTGTTTATAGCGCATTCGCAAGCTTCACTGACAGA 267
 Qy 103 HisAspTyrLeuArGAspGlyPheValSerArGThrMetGlyGluGluLeuProIlePro 122
 Db 268 CAAATCATTTTAAAAAATGTTTGGCGCAAAAAATCGTTCAAGAACTAAACCAAAA 327
 Qy 123 GlyLeuSerValLysTyrLeuArGlyLysValLysValThrAspLysValLysGluGln 142
 Db 328 TCTTTA-----TACGTTGCTTTTTCGTACGATGATGATGTTCTTAAAGAAAT 378
 Qy 143 LeuLysGlyIleThrAsp 148
 Db 379 TTGTCAAAAGTAAATCAT 396
 RESULT 8
 BH391233 721 bp DNA linear GSS 11-DEC-2001
 LOCUS AG-ND-141L2.TF ND-TAM Anopheles gambiae genomic clone AG-ND-141L2,
 DEFINITION DNA sequence.
 ACCESSION BH391233
 VERSION BH391233.1 GI:1737374
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
 ; Anophelinae.
 REFERENCE 1 (bases 1 to 721)
 AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Other_GSSs: AG-ND-141L2.TR
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.
 Seg primer: M13 For
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..721
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-141L2"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1, Site_1: HindIII"
 BASE COUNT 269 a 122 c 147 g 183 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.07e-10 Length: 721
 Score: 187.50 Matches: 62
 Percent Similarity: 43.10% Conservative: 38
 Best Local Similarity: 26.72% Mismatches: 108
 Query Match: 5.04% Indels: 24
 DB: 12 Gaps: 5

US-10-008-355-2 (1-712) x BH391233 (1-721)
 Qy 307 IleTyrPysGluAlaMetSerAlaSpGlnAlaThrArgIleLysTyrAlaSerLysTyr 326
 Db 23 GTATGAAACCAATATGATTAACACGATGCTACAAAGATTAGCTTACCGCTACATAT 82
 Qy 327 AlagInSerAlaAsnTyrTTrPlysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeu 346
 Db 83 GCAAGTGTGCTAACTAATTGCAAAAAACAGACCGCAACAATGAACGGGTTACAAAAAC 142
 Qy 347 AspValIleGlyArgLysArgAlaGluGluAlaArgAlaPheAlaAspTyrPheArgLysAsn 366
 Db 143 GTATCAATTGGTACAAAAAAGAAAGAACTGAGAAAAATACCAACAAATGGCGTAAATACCT 202
 Qy 367 GlyLysSerAlaValTyrGlyAspValLeuSerSerLeuGluLysAlaTyrLysGluLys 386
 Db 203 GAAACCAAGCTGTTTATGAAATATATAGCAATFACTGATGATCTACTACAGCAAAAT 262
 Qy 387 AlAlysAlaAsnArGlyLysThrTyrLeuSerGluThrLeuPheGlyLysThrGluVal 406
 Db 263 TCTACAGAAATATGAGAAAAATATATGTCACAGTTCACAGAAATGCTAAATATACATC 322
 Qy 407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr-----AsnProAspAla 422
 Db 323 AGAAATCTTTCACATGCGAGATGCTTAAACACGCTACATGACAGACGATGCATCTGCT 382
 Qy 423 HisAlaGlyIleLeuLysSerLeuAsp-----AspLysTyrLysAspTyrLeu 438
 Db 383 CAGGCACGTATGAACCTTAATTTACAGACGCTGTAACGACGCTTATGAAAGGTTTAAAT 442
 Qy 439 ProSerLeuAspArgLysValLeuPro-AlaMetLeuAspIleValArgArgArgIlePr 458
 Db 443 ACTCAACTTGAAGACGAATGCTATCTCAGATGCTTTGTTTATCAATCAAAAGGTGCG 502
 Qy 458 AlaLysPlysLeuProAspIlePheLys-----AsnValIle 470
 Db 503 AGCAATGATGCTGCTGCAACGTAAATCTGTAATCATCAGACGATCGCAATGCGATATGCG 562
 Qy 470 eAspLysLysPheLysGlyAspThrLysLysTyrAlaAspPheValIleAspLysSerVa 490
 Db 563 ACAGTCTCAATTCCTTGCAGACGACGCTTCACTAATTATCTCCGACCAATCCACAGCGC 622
 Qy 490 lValProTyrSerAsp-----LysPheHisAlaMetLeuLysSerMetAspLys 506
 Db 623 TGAATAATTAGCAATGATTAACCTTACACAGTTCCTGCACGCTATATCGAGACACAA 682
 Qy 506 s-----GlyLysPheAlaLysAlaIleGlu 514
 Db 683 AGTATTGGCTGAAAAATATGCTTAAATCTGATGAA 716
 RESULT 9
 BH394089 706 bp DNA linear GSS 11-DEC-2001
 LOCUS AG-ND-150P20.TF ND-TAM Anopheles gambiae genomic clone AG-ND-150P20
 DEFINITION , DNA sequence.
 ACCESSION BH394089
 VERSION BH394089.1 GI:17340230
 KEYWORDS GSS.
 SOURCE African malaria mosquito.
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
 ; Anophelinae.
 REFERENCE 1 (bases 1 to 706)
 AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
 JOURNAL Unpublished (2001)
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543

Email: b1loftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.

FEATURES
source
location/Qualifiers

1..706
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-150P20"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 267 a 122 c 140 g 177 t

ORIGIN

Alignment Scores:
Pred. No.: 4.96e-10 Length: 706
Score: 184.00 Matches: 49
Percent Similarity: 47.09% Conservative: 32
Best Local Similarity: 28.49% Mismatches: 82
Query Match: 4.95% Indels: 9
DB: 12 Gaps: 2

US-10-008-355-2 (1-712) x BH394089 (1-706)

QY 307 llettrpysglualametserralaasplnalatrarglielysTYrAlaSerlySTyr 326

DB 23 GTTATGAGAAAGCATATGATTAAGCATGCTACAGATAGCTTACGCGCTAATCAT 82

QY 327 lalaginserralaaspttlyrlypsasnserrilegilymetasargglyleualargleu 346

DB 83 GCAAGTGTGCTAACTATTGGAAAAAGCCGGAACAAATTGAAGCGGTTTACAAAAC 142

QY 347 AspValIleGlyArgLysArgLysArgLysArgLysArgLysArgLysArgLysArg 366

DB 143 GGAACAATTGTCGACAAAAAAGAACTGAGAAAAATATACCAACATGGCGCTAATAAGCT 202

QY 367 GlyLysSerAla-ValTYrGlyAspValLeuSerLeuGluLysAlaTYrLysGluG1 386

DB 203 GAAACAAAGCCGTGTATGAAATGTATTACCAATATCTGATCTACACCAATAT 262

QY 386 yAlaLysAlaAsnArgLysMetThrTYrLeuSerLysLysLysLysLysLysLysLys 406

DB 263 TTCTAACAGAAATATTCGAGAAAAATTATGTCACAGATTCACAGAAATGCTAATACAT 322

QY 406 lValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr-----AsnProAspAl 422

DB 323 CAGAAATTCCTTCGAGATCGGAGATGCTTAACAAAGCTACATGAAAGCATGCTATGTC 382

QY 422 ahisAlaGlyIleLeuLysSerLeuAsp-----AspLysTYrLysAspTYrLe 438

DB 383 TCAGGACGCTATGAAACCTAATTTAGAAAGACAGCTGTTAAGCAGGCTTATGAAGGGTTTAA 442

QY 438 uProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIleP 458

DB 443 TACTCACTTGAAGCAAGAAATGCTATCTCAGATGCTCTTATATATCATCAAAAGTATGC 502

QY 458 cAlaAspLysLeuProAspIlePheLysAsnVal 469

DB 503 AGCAATGTGCTCTCGCAACTGTAAATCTGTA 536

RESULT 10
BH400866 555 bp DNA linear GSS 11-DEC-2001
LOCUS BH400866/c
DEFINITION AG-ND-125M4.TR ND-TAM Anopheles gambiae genomic clone AG-ND-125M4,

DNA sequence.

ACCESSION
BH400866
VERSION
BH400866.1 GI:17347082

KEYWORDS
GSS.
African malaria mosquito.

SOURCE
ORGANISM
Anopheles gambiae

REFERENCE
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL
Unpublished (2001)

COMMENT
Other GSSs: AG-ND-125M4.TF

Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208
Fax: 301 838 3543

Email: b1loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 Rev
Class: BAC ends.

FEATURES
source
location/Qualifiers

1..555
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-125M4"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT 152 a 105 c 94 g 204 t

ORIGIN

Alignment Scores:
Pred. No.: 2.08e-09 Length: 555
Score: 177.00 Matches: 58
Percent Similarity: 46.82% Conservative: 23
Best Local Similarity: 33.53% Mismatches: 73
Query Match: 4.76% Indels: 21
DB: 12 Gaps: 4

US-10-008-355-2 (1-712) x BH400866 (1-555)

QY 310 GluAlaMetSerAlaAspGlnAlaThrArgLysTYrAlaSerLysTYrAlaGlnSer 329

DB 545 GAAATATGCGCTACTGATATATGCAACCCGATTAATATGCTTAATATGCAATCCGTCG 486

QY 330 AlaAsnTYrTrpLysAsnSerIleGlyMetAsnArgLysLeuAlaArgLeuAspValIle 349

DB 445 GCTAAGTATGCGAAGAAAAATATGATGCTAAGTA-GAAGGTTTAAAAATCTATAT-GCAGTA 428

QY 350 GlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsnGlyLysSer 369

DB 427 GCTAAAAAGCAGCCTTATGAAACCACTATGCA-----CAGAAAAACAGAGAGTA 377

QY 370 AlaValTYrGlyAspValLeuSerLeuGluLysAlaTYrLysGluGlyAlaLysAla 389

DB 376 AAAGCACTGTGATTAATTTAGCCAAITG-----TATACGAGCAGGCTCTTAT 326

QY 390 AsnArgLysMetThrTYrLeu-----SerGluThrLeuPheGlyGly 403

DB 325 GCATTGAACAATACCTTACTATCTAGGTTTACAAAAAATGCAGAGACATTTGGCTTGC 266

```

OY 404 ThrGIuValValArgPheAlaGlnPheAlaAsnAlaLeuAlaThrAspProAspAlaHis 423
Db 265 AACATTTTCATCATATATTTTCAGAGGTTGATCCGGAAGAAAGCTACCCCGGAATCTACT 206
OY 424 AlaGlyIleLeuLysSerLeuAspSplysTyrLysAspTyrLeuProSerLeuAspArg 443
Db 205 AAGAACAATGAAGAACACTTAACTCTTCTTCAAGAGATTTAAGAGAGAACTGATCT 146
OY 444 LysValIleuProAlaMetLeuAspIleValArgArgArgIleProAlaAspLysLeuPro 463
Db 145 AAGGTACAGCAAAATATATGCTTATATGCACAGAAAGCTCCGCTGAATTTTACT 86
OY 464 -----AspIlePheLysAsnVal 469
Db 85 TCAGATTGCTCAGTCTGATGTATAACAAACCTT 47

RESULT 11
LOCUS BH381877 769 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-155J17.TR ND-TAM Anopheles gambiae genomic clone AG-ND-155J17
ACCESSION BH381877
VERSION BH381877.1 GI:17328019
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
/ Anopheles.
REFERENCE 1 (bases 1 to 769)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other-GSSs: AG-ND-155J17.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
FEATURES
Location/Qualifiers
source 1..769
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-155J17"
/clone_lib="ND-TAM"
/note="Vector: pECBAC1, Site_1: HindIII"

BASE COUNT 286 a 124 c 156 g 203 t
ORIGIN

Alignment Scores:
Pred. No.: 3,46e-09 Length: 769
Score: 177.00 Matches: 63
Percent Similarity: 41.84% Conservative: 37
Best Local Similarity: 26,36% Mismatches: 87
Query Match: 4,76% Indels: 53
DB: 12 Gaps: 6

US-10-008-355-2 (1-712) x BH381877 (1-769)

```

```

OY 307 IleTyrPlyGluAlaMetSerAlaAspGlnAlaThrArgIleLysTyrAlaSerLysTyr 326
Db 21 GTTATGAGAGACATATGATTAAGACGATGCTCAAGATTAAGCTTACCGCTTAACAT 80
OY 327 AlaGlnSerAlaAsnTyrTrpLysAsnSerIleGlyMetAsnArgGlyLeuAlaArgLeu 346
Db 81 GCAGGTGTGCTAACTATTGAAAAAGAGACCCGACAAATTTGAACGCGTTTACAAAAAC 140
OY 347 AspValIleGlyArgLysArgAlaGluGluArgAlaPheAlaAspTrpIleArgLysAsn 366
Db 141 GGAACAATGTGTGACAAAAAAGAGTTGAGAAAAATTTTAA-TGGCGCTGATTAAGCT 199
OY 367 GlyLysSerAlaValTyrGlyAspValLeuSerLeuGluLysAlaTyrLysGluGly 386
Db 200 GAAAAACAAGCTGTTTATGAAATGATATTACAAATACTGATGACTACAAACCAATT 259
OY 387 AlaLysAlaAsnArgGluMetThrTyrLeuSerGluTrpLeuPheGlyGlyThrGluVal 406
Db 260 TCTAACAGAAATATCGAGAAAAATTTATGCTGCACAGTTCCAGAGAAATGCTTAATACATC 319
OY 407 ValArgPheAlaGlnPheAlaAsnAlaLeuAlaThr-----AsnProAspAla 422
Db 320 AGAATTCTTCCAGATCGGAGATGCTTACCAAGCTACATGAAACAGATGATCTGCT 379
OY 423 HisAlaGlyIleLeuLysSerLeuAsp-----AspLysTyrLysAspTyrLeu 438
Db 380 CAGCAGCTATGAACAACTAAATTAGAACACGCTGTAAAGCGCTTATGAAGGCTTAAT 439
OY 439 ProSerLeuAspArgLysValLeuProAlaMetLeuAspIleValArgArgArgIlePro 458
Db 440 ACTCAACTGAGCAGAAATGCTATCTCAGATGGCTTCTTATATCAATCAAAATGACA 499
OY 459 AlaAsp-----LysLeuProAspIlePhe 466
Db 500 GCAGATGTGCTGCTTGTGCAACTGTAAATCTGTAATGCATGACAACTTGGCAATATGCA 559
OY 467 LysAsnValIle----- 470
Db 560 CAGCTTCAATTTTTCGAAAGCAGCTTCAGTAATTAATCTTCGAAACAAATCCAGCGCT 619
OY 471 -----AspLysPheLys-----GlyAspThrLys 479
Db 620 GAAAAATTAGCATGATGATTAACCTTACAAAGTTTCTGACGCTATATCGGAGACACAAG 679
OY 480 LysTyr-----AlaAspPheValPheAspLysSerValIleProTyr 493
Db 680 TATGCTGAAAAATATGCTAAACTGATGAGGTTTCAGAAAGACAGCGCTTATTAT 736

RESULT 12
LOCUS BH164253/c 1005 bp DNA linear GSS 24-SEP-2001
DEFINITION ENT163Tf Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH164253
VERSION BH164253.1 GI:15737691
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 1005)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library

```



```

FEATURES
Source      Location/Qualifiers
1.2349      /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="MGD:1895443"
            /db_xref="taxon:10090"
            /clone="2810401M23"
            /dev_stage="10, 11 days embryo"
            /note="data source:SPTR, source key:Q9NYD8, evidence:ISS
            homolog to CGI-201 PROTEIN
            putative"
CDS
            /codon_start=1
            /protein_id="BAB28572.1"
            /db_xref="GI:12850038"
            /translation="MAASTAGKQRIPIKVAKNKPAEVOITAEQLREARELEL
            LPPEQKITDEEELNDYKLRKRKFEENIKRNVISNVIKQWESLEIQRAS
            IYRALDVIDYENITLMLKYAMEMKRNQVNAHNMORAITTLPVNOFWYKTYMEE
            MLGNVAGROYFEREMWEPEDQAMHSYINELRYKKEVERARTYERVLVHPAKNM
            IKYARPEKHAIFAHARKYERAVEFDEDMEDLYAFAPKEBNQEFERVITYK
            YALDRISKQDADELKNTITEKFGDRGLIEDITVSRKQYEEVAPNPDAMF
            DYLRLVESDAADAVREYERAIANVPIQERHMKRYLYLVNVALYEADPER
            TROYOASLELIPHKFTFAKMLLYAFAEIRKRLPARALGTSIGCKPKLFG
            YIELEQLREDRCKLYEKFLERGPENCTSMIKPAELETIGIERARAYELAISO
            PRDMPEVLYEANKTRNCEKEERLMLESRSFDEGTYSDEKRVKLPMEKVKKR
            KRVQADDSGDAEMETIDYIFEDANQPNKLAMATKMKQDEREADADDPKDI
            DESESSP"
BASE COUNT      744 a      451 c      597 g      557 t
ORIGIN
Alignment Scores:
Pred. No.:      1.27      Length:      2349
Score:          107.50      Matches:      115
Percent Similarity: 34.23%      Conservative: 89
Best Local Similarity: 19.30%      Mismatches: 203
Query Match:    2.89%      Indels:      189
DB:            11      Gaps:      32
US-10-008-355-2 (1-712) x AK012962 (1-2349)
QY      62 IleaAsnAlaValAlIlePheGlyGlyGlyCys---ThrGlyIleThrValSerAsp 80
      19 GGGCGTGAAGTCTGTGTGGCGCGCGTTCGAGTCTTTCCAGTCCGCGTTGTGCGCC 78
QY      81 Gln-GlyIleuIlePheThrAsnHisHisCysGlyTyrglyAlaIleGlnSerGlnSerTh 100
      79 AGGATCGCTCGAGTTC-----CACGGCTCGAGTCAATGCGACGCTCCAGCGC----- 124
QY      100 rValAspHisAspTyrglyLeuArgAspGlyPheValSerArgThrMetGlyGlnGluLeuPr 120
      125 -----GGCGCGGAACAGC 138
QY      120 o-IleProGlyLeuSerValIlyTyrglyLeuArgLysIleValIlyValIThrAspLysValG 140
      139 GCATTCT-----AAAGTGGCCAAAGGTGAAGAACAGAGCGCC 174
QY      140 IugIyGlnLeuLysGlyIleThrAspGluMetGluArgLeuArgLysAlaGlnGluValC 160
      175 CGCGCGAAGTGCAG---ATAACGGCCGAG---CAGCTCTCCGAGAGGTAAGAAGAGAG 228
QY      160 ysgGlnGluLeu-----AlaLysLysGlnLysAlaAspGluAsnGlnLeuCysI 176
      229 AACTCGAGCTCTCCACCTCCACCTCAGCAGAGATCAGATGAGGGAAGAAATTA----- 284
QY      176 IeValGluProPheTyrrSerAsnAsnGluTyrrPheLeuIleValTyrrAspValPheLys 196
      285 -----AACGATTACAAACTCCGGAAGAAAGAACTTTTGAAG 321
QY      196 sPValArgMetValPheAlaProPheSerValGlyLysPheGlyGlyAspThrAspA 216
      322 AT-----ACATTAAGAAAAAATAGAGACTGTGATTAGCA 354

```

```

Oy 550 hepAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAspAlaAsnPheT 570
Db 1212 -----TGAAGCGCTACATCTAC----- 1229
Oy 570 hMetArgMetSerTyrGlySerIleLeuGlyTyrGluProGlnAspGlyAlaTTPtyrA 590
Db 1230 -----CTTGGGTTA 1239
Oy 590 sNTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheA 610
Db 1240 ACTATGCACTCTATGAAG-----CTGAAGCCACAGATCCCGACAGACACAGACAG 1293
Oy 610 laValGlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyr 624
Db 1294 TATATCAAGCGTCT--TTGAACTAATTCTCACCAAAAGTTTC 1334

```

Search completed: October 18, 2002, 00:00:54
 Job time : 1734 secs